THE UNIVERSITY *of* EDINBURGH

The Epidemiology, Classification and Evolution of Human Rhinoviruses

Chloé Leanne McIntyre

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

March 2013

Abstract

Human rhinoviruses (HRV) are extremely common human respiratory pathogens, most commonly associated with mild upper respiratory tract infections. The three known species of HRV (HRV-A, -B and –C) are members of the family *Picornaviridae* and genus *Enterovirus*. In contrast to the enterovirus (EV-A-D) species that commonly infect the gut, HRV are generally thought to be acid labile with replication restricted to the respiratory tract.

Investigations of the clinical correlations of HRV infections detected on diagnostic screening of respiratory specimens demonstrated no specific association between HRV variant and clinical presentation. For example, similar species distributions were observed in patients admitted to the ITU and those discharged with minor illness. Unexpectedly, screening of stool specimens for HRV showed a prevalence of 10% with viral loads similar to EV infections. These findings suggested that a reappraisal of HRV tropism and disease associations may be warranted.

HRV-A and -B isolates were originally classified into 100 serotypes by serological neutralisation properties. As HRV-C is difficult to isolate, no attempt had been previously made to classify the wealth of available HRV-C sequences. To facilitate definition of novel HRV types and classification of HRV-C, a system was devised to divide HRV sequences into genotypically defined types. Pairwise VP1 nucleotide p-distance analysis revealed distinct thresholds between inter- and intra- type divergence and available sequences were classified into 77 HRV-A, 29 -B and 51 -C types. This provides a standardised basis for type definition and identification, allowing consistency in studies of genetic diversity, epidemiology and evolution. It has been adopted by the ICTV Picornavirus Study Group for classification of HRV.

Although the occurrence of recombination has been documented within the coding region of EV, analysis of dated HRV sequences revealed an overall lack of intra-species recombination between three coding regions of HRV-B and -C. In contrast, full HRV-A type groups appeared to have been subject to a large number of recombination events, suggesting extensive recombination during the period of its diversification into types. Putative recombination breakpoints localised to the non-structural region. Within HRV-A and HRV-B, recombination within the 5'UTR was infrequent. However, over 60% of analysed HRV-C strains grouped within the HRV-A clade and two recombination hotspots were identified. An additional interspecies recombination event was detected between HRV-A/C in the 2A coding region, with putative breakpoints mapping to the boundaries of the C-terminal domain of the proteinase.

The studies within this thesis provide evidence for a broadened understanding of the clinical significance of HRV. In addition, the assignment of HRV sequences into genotypically defined types allowed description of the observed genetic diversity and completion of analysis which reaffirmed the sporadic nature of recombination within the coding region of HRV.

Declaration of Originality

I declare that the work recorded within this thesis was completed entirely by myself at the Centre for Infectious Diseases and Roslin Institute of the University of Edinburgh between September 2009 and March 2013. In cases where others have made a contribution to the results obtained, this is clearly stated within the text. This work has not been submitted for any other degree or professional qualification. In cases where portions of the work contained within this thesis have been published in academic journals, this is indicated within the text and copies of papers are given as appendices. Where required by license agreement, permission to include published papers has been sought from publishers.

Chloé McIntyre

Acknowledgements

A hundred people have contributed in a thousand unique ways to the completion of this PhD and there is not enough paper in the world to thank them all as much as they deserve.

For my supervisor, Professor Peter Simmonds. Thank you for giving this opportunity to a young surgical trainee with little experience. Your unending patience, enthusiasm and dedication over the last 4 years has taken me from a clueless newcomer to a fledgling scientist able and willing to think for herself. I cannot thank you enough.

For the others who have been involved in my academic and scientific growth, a special thanks. For Professor Paul Sharp who has always been available for advice and support. For Dr Heli Harvala, whose infectious enthusiasm and incredible passion for virology has widened the horizons of my interest and experience far beyond what I initially imagined. For Dr Carol McWilliam-Leitch who took me from having never held a pipette to confidently designing my own PCR protocols far faster than I ever thought possible. For all the PhD students and post-docs in our lab past and present; Elly, Ingrid, Sinead, Nora, Nigel, Jereon, Donald, Jill, Colin, Inga, Nicky, Richard and Carol. Not only have you all been at various times an indispensible source of wisdom, but you have brightened my days and almost never been too busy to meet for a beer and a giggle after hours. To the students whose projects I have been fortunate enough to be involved in, thank you for allowing me to learn how to teach and keeping me smiling with your enthusiasm and humour.

For my wonderful parents who have been unfailingly loving and supportive and never questioned my decision to follow my passion. Without you, none of this would have ever been possible. You have always made me feel like a star! I cannot thank you enough.

For my best friend, life-partner, handyman, life-coach, chef, therapist, IT-support, removal man, musical hero, brew-master and most of all, amazing husband; Kevin. Thank you for being so endlessly, uncomplainingly, beautifully patient with me; with my long working hours, tiredness, frustration and endless, obsessive fixation on minute details. Your love and support has been the cornerstone of everything I've done. And most of all, to our little bun in the oven. I hope that someday you will read this and smile.

Mummy loves science and Mummy loves you.

List of Original Publications

Original publications directly associated with this thesis

McIntyre CL, McWilliam Leitch EC, Savolainen-Kopra C, Hovi T, Simmonds P. 2010. Analysis of genetic diversity and sites of recombination in human rhinovirus species C. J Virol **84**:10297–310.

McIntyre CL, Savolainen-Kopra C, Hovi T, Simmonds P. 2013. Recombination in the evolution of human rhinovirus genomes. Arch Virol. [Epub ahead of print]

Simmonds P, McIntyre CL, Savolainen-Kopra C, Tapparel C, Mackay IM, Hovi T. 2010. Proposals for the classification of human rhinovirus species C into genotypically assigned types. J Gen Virol **91**:2409–19.

Harvala H, McIntyre CL, McLeish NJ, Kondracka J, Palmer J, Molyneaux P, Gunson R, Bennett S, Templeton K, Simmonds P. 2012. High detection frequency and viral loads of human rhinovirus species A to C in fecal samples; diagnostic and clinical implications. J Med Virol **84**:536–42.

McIntyre CL, Knowles NJ, Simmonds P. 2013. Proposals for the classification of human rhinovirus species A, B and C (HRV-A, -B and –C) into genotypically assigned types. *Submitted.*

Original publications to which the author has otherwise contributed

Gaunt ER, Harvala H, McIntyre CL, Templeton KE, Simmonds P. 2011. Disease burden of the most commonly detected respiratory viruses in hospitalized patients calculated using the disability adjusted life year (DALY) model. J Clin Virol. **52**(3):215-21

McLeish NJ, Witteveldt J, Clasper L, McIntyre CL, McWilliam Leitch EC, Hardie A, Bennett S, Gunson R, Carman WF, Feeney SA, Coyle PV, Vipond B, Muir P, Benschop K, Wolthers K, Waris M, Osterback R, Johannessen I, Templeton K, Harvala H, Simmonds P. 2012. Development and assay of RNA transcripts of enterovirus species A to D, rhinovirus species A to C, and human parechovirus: assessment of assay sensitivity and specificity of real-time screening and typing methods. J Clin Micro **50**:2910–7.

Harvala H, Gaunt E, McIntyre CL, Roddie H, Labonte S, Curran E, Othieno R, Simmonds P, Bremner J. 2012. Epidemiology and clinical characteristics of parainfluenza virus 3 outbreak in a Haemato-oncology unit. J Infect. **65**(3):246-54

Harvala H, McIntyre CL, Imai N, Clasper L, Djoko CF, LeBreton M, Vermeulen M, Saville A, Mutapi F, Tamoufé U, Kiyang J, Biblia TG, Midzi N, Mduluza T, Pépin J, Njouom R, Smura T, Fair JN, Wolfe ND, Roivainen M, Simmonds P. 2012. High seroprevalence of enterovirus infections in apes and old world monkeys. Emerg Infect Dis **18**(2):283–6.

Harvala H, Mcleish N, Kondracka J, Mcintyre CL, Leitch ECM, Templeton K, Simmonds P. 2011. Comparison of Human Parechovirus and Enterovirus Detection Frequencies in Cerebrospinal Fluid Samples Collected Over a 5-Year Period in Edinburgh : HPeV Type 3 Identified as the Most Common Picornavirus Type. J Clin Micro **896**:889–896.

Abbreviations

AdV	adenovirus
A&E	Accident and Emergency department
AICc	small sample Akaike information criterion
AOM	acute otitis media
Arg	arginine
ATCC	American Tissue Culture Collection
BEAST	Bayesian Evolutionary Analysis of Sampling Trees
BEAUti	Bayesian Evolutionary Analysis Utility
BLAST	Basic Local Alignment Search Tool
BRAV	Bovine rhinitis A virus
BRBV	Bovine rhinitis B virus
CAV	Coxsackie A virus
CBV	Coxsackie B virus
cDNA	complementary deoxyribonucleic acid
CFLI	cold and flu-like illness
CHI	community health index
CNS	central nervous system
cre	*cis-acting* replication element
CSF	cerebrospinal fluid
Cys	cysteine
CT	cycle threshold for real-time PCR
DAF	decay accelerating factor
DDBJ	DNA Databank of Japan
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
dsRNA	double stranded ribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
EMCV	encephalomyocarditis virus
ERAV	equine rhinitis A virus
ERBV	equine rhinitis B virus
EV	enterovirus

FluA	influenza A virus
FluB	influenza B virus
FMDV	foot and mouth disease virus
GARD	genetic algorithm for recombination detection
GP	general practitioner
HAV	hepatitis A virus
HCoV	human coronavirus
HCV	hepatitis C virus
HDU	High Dependency Unit
HIV	human immunodeficiency virus
HKY	Hasegawa-Kishino-Yano model of nucleotide substitution
HPA	Health Protection Agency
HMPV	human metapneumovirus
HRV	human rhinovirus
HSV	herpes simplex virus
ICAM-1	intercellular adhesion molecule 1
ICTV	International Committee on the Taxonomy of Viruses
IFA	immunofluorescence
indels	insertions and deletions in nucleotide sequences
IRES	internal ribosomal entry site
ITU	Intensive Care/Therapy Unit
kB	kilobase
L	leader protein
LDL	low density lipoprotein
LRTI	lower respiratory tract infection
MAU	Medical Assessment Unit
MCL	maximum composite likelihood
MCMC	Markov Chain Monte Carlo
MEGA	Molecular Evolutionary Genetics Analysis
NCBI	National Center for Biotechnology Information
NJ	neighbour joining
NPEV	non polio enterovirus
nt	nucleotide
OTU	operational taxonomic unit
pat	provisionally assigned human rhinovirus type

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PIM2	Paediatric Index of Mortality score
PIV	parainfluenza virus
PPT	polypyrimidine tract
PV	poliovirus
RDP	Recombination detection program
RF	recombinant form
RHSC	Royal Hospital for Sick Children (Edinburgh)
RIE	Royal Infirmary of Edinburgh
RNA	ribonucleic acid
RT	reverse transcription
RT-PCR	reverse transcription polymerase chain reaction
SBP	Single breakpoint recombination detection
SGPB	*Streptomyces griseus* proteinase B
SVC	Scottish Virology Centre
TAE	tris-acetate EDTA
Tm	melting temperature
UTR	untranslated region
URTI	upper respiratory tract infection
UV	ultraviolet
VPg	viral protein genome linked
VZV	varicella zoster virus
WHO	World Health Organisation

List of Figures

List of Tables

Contents

Chapter 1
Introduction

1.1 Introduction

Human rhinoviruses (family *Picornaviridae*, genus *Enterovirus*) are best known as one of the major causes of the common cold. Although mainly associated with a mild and self-limiting upper respiratory tract infection (URTI), human rhinoviruses (HRV) are an established cause of other more serious clinical illness (Broberg et al., 2011; Hicks et al., 2006; Gutman et al., 2007; Fuji et al., 2011; Tapparel et al., 2009c). Childhood wheezing secondary to HRV infections has also been extensively implicated in the subsequent development of asthma (Jartti and Korppi, 2011; Jackson et al., 2008; Gern and Busse, 1999). In addition, the common cold places a considerable economic and social burden upon suffering populations, in terms of lost productivity, purchase of over-the-counter remedies and even unnecessary prescription of antibiotics (Fendrick et al., 2003). The last ten years have seen HRV research blossom with the discovery of an entire new species, many new types and ever increasing association with more severe disease. As the scope of clinical impact of HRV becomes clear, the necessity of developing an understanding of the evolutionary processes shaping the enormous diversity of these viruses and developing formal criteria for their classification has become evident. The studies presented within this thesis have aimed to address aspects of these issues.

1.2 Taxonomy and classification of viruses

1.2.1 General virus classification

The taxonomic classification and naming of viruses in a manner which allows worldwide communication between researchers has long been an important issue.

Since 1971, a broad method of virus classification related to the nature of the viral genome and the method of translation has been used for high-level grouping of viruses. This is known as the Baltimore Classification System (Baltimore, 1971) and divides viruses into seven groups based on nucleic acid type, sense of genome and method of replication. Although this system remains extremely useful for developing a broad understanding of the properties of certain virus groups, the vast number of diverse subgroups within each class highlights the need for a system of defining and cataloguing smaller groups.

The International Committee for the Taxonomy of Viruses (ICTV) has overseen issues relating to the naming and taxonomic grouping of viruses since the 1970s. The committee exists to provide a universal system of classification and nomenclature and releases in the form of reports and amendments are the state of the art in the field. The currently used system of virus classification is similar to, but distinct from other biological classification systems. Classification is based on sequence data and certain biological properties where known. Groups of virus variants, strains or types are classified as species. The ICTV defines a virus species as "a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria" (www.ictvonline.org). Distinguishing features of different species include phylogenetic relationships, sequence identity, host range, cell tropism, pathogenesis and antigenicity. The assignment and naming of variants, strains or types is overseen by various subcommittees or study groups with a particular expertise in a certain virus family. In ascending order of magnitude, virus species are grouped together according to common aspects and phylogenetic relationships to form a genus, a group of genera forms a family and a group of families forms an order. In instances of complex hierarchical problems, subfamilies of several genera can be created. However, subfamily membership is not a necessary aspect of a particular virus' taxonomy. There are currently, as of 2012, 7 recognised

orders, 96 families, 22 subfamilies, 420 genera and 2618 species of virus[1]. Although all proposals for naming and taxonomy of viruses above the species level must be ratified by ICTV members, each virus family is overseen by a subcommittee or study group who are responsible for the development and submission of such proposals.

1.2.2 Classification of the *Picornaviridae*

Picornaviruses are one of five virus families which fall within the order *Picornavirales* (also including *Dicistroviridae, Marnaviridae, Iflavirus* and *Secoviridae)* (Sanfaçon et al., 2011). The order consists of a diverse range of viruses with a single stranded positive sense ribonucleic acid (RNA) genome, grouped together on the basis of similar capsid structure and viral life cycles. *Picornavirales* have a very wide host range including humans, animals, plants, algae, insects and birds. In terms of viral capsid structure, these diverse viruses share a non-enveloped capsid consisting of 60 protomers, each with three distinct "jelly-roll" domains (Le Gall et al., 2008). The formal designation of the order *Picornavirales* replaces the putative "picorna-like supergroup". In addition to the five known families within the order, there are two unassigned genera (*Bacillarnavirus* and *Labyrnavirus*) which belong to no known family (www.picornavirales.com).

The largest family within the order *Picornavirales* are the *Picornaviridae*. Not only is this an extremely large and genetically diverse virus family, but it contains many human and veterinary pathogens of great clinical importance, including poliovirus (PV) and foot and mouth disease virus (FMDV). The family *Picornaviridae* currently contains 17 recognised genera (Knowles et al., 2012)[2] which cause clinical

[1] These figures have been updated since the publication of the Ninth Report of the ICTV and the figures reported here reflect these updates (www.ictvonline.org).

[2] Five of the seventeen picornavirus genera mentioned (Aquamavirus, Cosavirus, Dicipivirus, Megrivirus and Salivirus) were approved in Feburary 2013 and hence did not appear in the 9th ICTV report released in 2012. These newly approved genera appear in the revised ICTV Master Species List (www.ictvonline.org). In addition, over 20 currently unassigned putative picornaviruses have been found in such diverse host species as bats, eels, cats, chickens, mice and tortoises (www.picornaviridae.com).

disease in a wide range of human and animal hosts (Table 1.1). The family is named for the fact that all members are extremely small (pico) RNA viruses.

TABLE 1.1: Members of the family Picornaviridae and their natural hosts

Genera	**Species**	**Number of types**[a]	**Host**
Apthovirus	Foot and mouth disease virus (FMDV)	7	Cloven-footed animals (eg/ cattle and pigs)
	Bovine rhinitis A virus (BRAV)	2	Cows
	Bovine rhinitis B virus (BRBV)	1	Cows
	Equine rhinitis A virus (ERAV)	1	Horses
Cardiovirus	Encephalomyocarditis virus (EMCV)	1	Pigs
	Theilovirus	12	Rodents Humans
Aquamavirus	Aquamavirus A	2	Seals
Avihepatovirus	Duck hepatitis A virus	3	Ducks
Cosavirus	Cosavirus A	24	Humans
Dicipivirus	Cadicivirus A	1	Dogs
Enterovirus	Enterovirus A (EV-A)	23	Humans Primates
	Enterovirus B (EV-B)	60	Humans Primates Pigs
	Enterovirus C (EV-C)	23	Humans
	Enterovirus D (EV-D)	5	Humans Primates
	Enterovirus E (EV-E)	4	Cows
	Enterovirus F (EV-F)	6	Cows Possums
	Enterovirus G (EV-G)	6	Pigs Sheep
	Enterovirus H (EV-H)	1	Primates
	Enterovirus J (EV-J)	6	Primates
	Human rhinovirus A (HRV-A)	77[b]	Humans
	Human rhinovirus B (HRV-B)	25[b]	Humans
	Human rhinovirus C (HRV-C)	51[b]	Humans
Erbovirus	Equine rhinitis B virus (ERBV)	3	Horses
Hepatovirus	Hepatitis A virus	1	Humans

Kobuvirus	Aichivirus A	3	Humans Dogs Mice
	Aichivirus B	2	Cows
	Aichivirus C	1	Pigs
Megrivirus	Melegrivirus A	1	Turkeys
Parechovirus	Human parechovirus	16	Humans
	Ljungan virus	4	Rodents
Salivirus	Salivirus A	2	Humans Primates
Sapelovirus	Porcine sapelovirus	1	Pigs
	Simian sapelovirus	1	Primates
	Avian sapelovirus	1	Birds
Senecavirus	Seneca valley virus	1	Pigs
Teschovirus	Porcine teschovirus	13	Pigs
Tremovirus	Avian encephalomyelitis virus	1	Birds

[a] Information included on number of types and known host species is adapted from information available on the Picornavirus study group website (www.picornaviridae.com) and additional references (Pallansch and Roos, 2007; Knowles et al., 2012).

[b] The number of HRV types given relates to the number of types officially recognised in early 2013. Studies presented within this thesis have identified and proposed a number of additional HRV types from all three species.

The criteria for the definition of members of the same genus within the *Picornaviridae* are homology between certain important polypeptides (specifically, the leader protein, 2A, 2B and 3A coding regions), RNA structural homology of the internal ribosomal entry site (IRES) and amino acid identity of more than 40% in the P1 and P2 coding regions and more than 50% in the P3 coding region (Knowles et al., 2012).

1.2.3 Classification of the genus *Enterovirus*

The enteroviruses (EV) originally consisted of only viruses isolated from humans and were divided into four main groups based on their pathogenesis in man and suckling mice. The Coxsackie A viruses (CAV) were first isolated in 1948 from stool samples of children with an illness clinically similar to poliomyelitis (Dalldorf and Sickles, 1948). In contrast to PV which infected primates, CAV caused a flaccid paralysis in suckling mice. Coxsackie B viruses (CBV) were isolated in 1949 from

cases of aseptic meningitis (Melnick et al., 1949) and were found to cause a spastic paralysis and severe systemic infection in suckling mice. The final classical group of human EV were the echoviruses. These were first isolated in 1951 (Robbins et al., 1951) from tissue culture samples. They are named enteric cytopathogenic human orphan viruses due to their initial lack of association with clinical disease in either humans or laboratory animals.

While the classification based on pathogenicity was useful, as sequence data accumulated it became apparent that these subgroups did not correlate with observed phylogenetic relationships (Hyypiä et al., 1997). In addition, each observed type appeared to cause a wide spectrum of disease and so classification in clinical terms was impossible. Therefore, human EV were classified into four species (EV-A – D) on the basis of sequence identity and phylogenetic relationships. Within each species, individual types were defined either by investigation of serological properties or sequence identity (discussed in Chapter 4).

The classification of picornaviruses on the basis of phylogenetic relationships and sequence identity has led to the recognition that a great many animal viruses additionally fall within the *Enterovirus* genus (Table 1.1). Indeed, many EV that infect primates are members of the same species as those infecting humans. For instance, EV-A, -B and –D all predominantly infect humans and contain several simian EV (Harvala et al., 2011; Oberste et al., 2002; Harvala et al., 2012a). The recognition of this host species diversity within a single EV species has resulted in the recent proposal to remove the host species from the enterovirus name (resulting in nine EV species named EV-A – J[3]). HRV (for which no other host species is currently known) were also reclassified as members of the *Enterovirus* genus on the basis of observed similarities in life cycle, genome organisation and phylogenetic relationships (Laine et al., 2005)

[3] EV-I was excluded from the naming system in order to avoid confusion with EV-1.

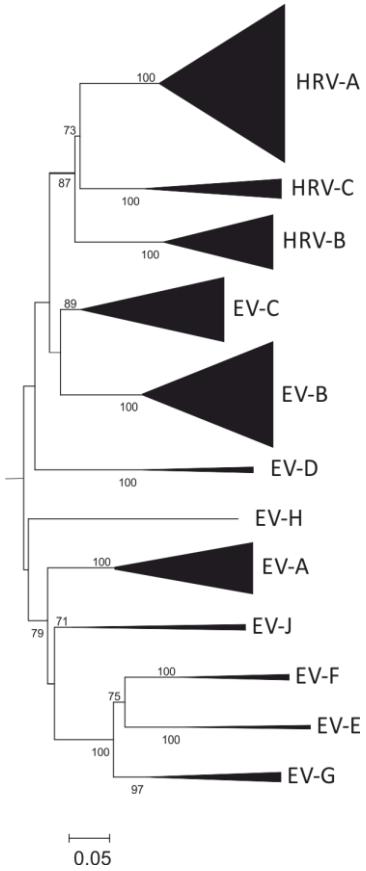


FIGURE 1.1: Neighbour joining phylogenetic tree depicting relationships between species within the Enterovirus genus. *Tree was constructed as described in Chapter 2 and using simian sapelovirus sequence AY064708 as an outgroup. Branch to tree root has been collapsed. Enterovirus species have been collapsed for ease of reference. Host species for each EV species are given in Table 1.1. Branches are scaled by genetic distance.*

1.2.4 Classification of human rhinoviruses (HRV)

1.2.4.1 Classification of human rhinoviruses based on biological and serological properties

Since the isolation of HRV-A1 in 1956 by two independent laboratories (Pelon et al., 1957; Price, 1956), the classification of HRV has undergone many revisions. Early studies of the biology of HRV sought to classify them as "H" or "M" strains according to their ability to grow in various different cell lines (Andrewes, 1961). 29

strains of HRV were classified in this way, with the vast majority being found to replicate only in human embryonic kidney cells (designated H strains). Only 2 HRV strains (which are known under the current classification system as HRV-A1 and HRV-A2) were found to be able to replicate in both human and monkey kidney cells (designated M strains) (Taylor-Robinson and Tyrrell, 1962; Arden and Mackay, 2010). This particular classification of HRV based on properties of *in vitro* cell culture was abandoned and is now of historical interest only.

Between 1967 and 1987, all of the known HRV isolates were collected and tested by serological neutralisation assays to determine their antigenic relationships (Kapikian et al., 1967; Hamparian et al., 1987; Kapikian et al., 1971) This collaborative program was completed in three stages and resulted in the definition of 100 sequentially numbered HRV serotypes. As the existence of distinct species of HRV was not recognised at this time, sequential numbering remains interspersed throughout the two HRV species now known as HRV-A and HRV-B[4]. Serotypes were initially defined on the basis of antigenic distinctness and a lack of significant cross-reactivity with known HRV serotypes (Kapikian et al., 1967). However, subsequent studies identified a number of HRV serotype pairs which exhibited reciprocal cross-reactivity (HRV-1A/1B, 2/49, 3/14, 9/32, 12/78, 13/41, 15/74, 29/44, 36/58, 8/95) and an even greater number of serotype pairs which exhibited non-reciprocal cross-neutralisation (Cooney et al., 1982; Ledford et al., 2004; Halfpap and Cooney, 1983). However, the antibody titres to heterotypic HRV were generally low.

An additional classification criteria applied to HRV-A and –B was cell surface receptor usage. Before the major receptors for HRV infection were known, competitive binding assays were used to demonstrate the existence of two distinct receptor groups. If a cell culture assay was saturated with one HRV type, most other

[4] For clarity, HRV types are now conventionally listed with their species designation included (for example, HRV-A12 and HRV-B3).

known types could not bind and establish infection. In fact, only 4 of 24 HRV types showed no inhibition. It was therefore inferred that the majority of HRV used the same receptor (Abraham and Colonno, 1984). This "major group" HRV receptor was eventually identified as the cell surface glycoprotein intercellular adhesion molecule 1 (ICAM-1) (Greve et al., 1989); a protein which is expressed in abundance on endothelial cells. All known HRV-A and –B were then classified into two groups on the basis of receptor usage (Uncapher et al., 1991); with 91 serotypes falling within the major group and 9 within the minor. The minor group receptor was shown to be a member of the low density lipoprotein family (LDL) (Hofer et al., 1994). One type, HRV-87, was determined to use a third receptor (Uncapher et al., 1991). However, this was later reclassified as a member of the EV-D species (EV-68) (Blomqvist et al., 2002b; Ishiko et al., 2002) with tropism for the respiratory tract.

Finally, HRV have also been classified in terms of sensitivity to certain classes of antiviral drugs. In 1990, all 100 HRV serotypes were tested against a specific panel of 15 antiviral compounds and this revealed the existence of two antiviral susceptibility groups (Andries et al., 1990). Group A consisted of 33 serotypes which were susceptible to elongated compounds, such as WIN517111 and Group B consisted of 67 serotypes susceptible to short compounds, such as chalcone, dichloroflavan and R61837. Further analysis of the amino acid sequence of this region revealed that specific mutations were predictive of the susceptibility of a particular HRV serotype to a further antiviral drug, known as pleconaril (Ledford et al., 2005).

Although methods of HRV classification based on observable biological properties have proved extremely useful in HRV research, all require extensive and time-consuming laboratory procedures. The use of phylogenetic methods based on readily

available sequence data for the classification of HRV has allowed much more timely and easy identification.

1.2.4.2 Classification of human rhinoviruses based on phylogenetic relationships

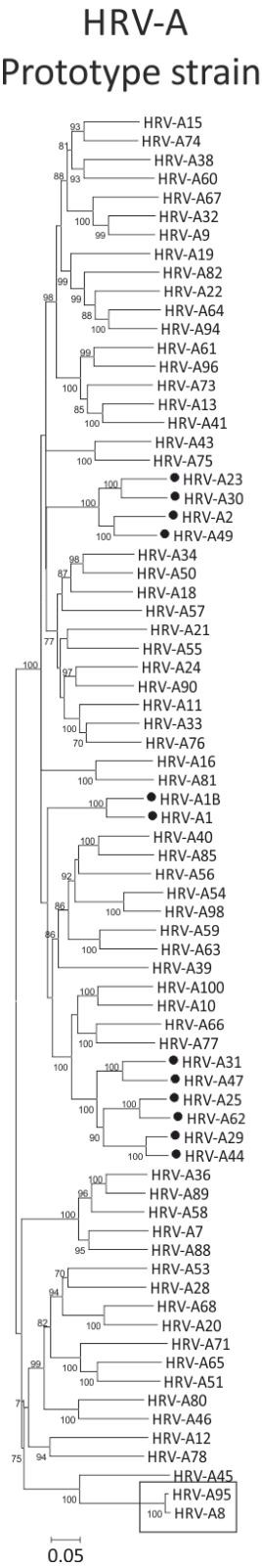
The recognition that the 100 recognised serotypes of HRV could be divided into two species (HRV-A and –B) was achieved by phylogenetic analysis. These two distinct species groupings were shown to be present throughout the capsid coding region (Horsnell et al., 1995; Savolainen et al., 2004; Laine et al., 2005), the 5' untranslated region (UTR) (Kiang et al., 2008) and the 3Dpol region at the 3' end of the coding genome (Savolainen et al., 2004). The completion of full genome sequences for all 100 HRV-A and –B strains in 2009 additionally confirmed these two species groupings (Palmenberg et al., 2009).

The common usage of molecular methods in studies of HRV epidemiology has led to the rapid accumulation of sequence data and a revised understanding of many aspects of HRV classification and taxonomy. As previously mentioned, HRV were initially classified within the genus *Rhinovirus* (distinct from *Enterovirus*) on the basis of cell tropism and disease manifestations. HRV generally replicate in the respiratory tract and cause respiratory disease, whereas EV predominantly replicate in the gastrointestinal tract. However, studies of phylogenetic relatedness and sequence identity between these two groups has led to the re-classification of HRV within the genus *Enterovirus* (Laine et al., 2005; Savolainen et al., 2004). It has been shown that the division of HRV into subgroups based on biological properties, as described above, does not correlate exactly with the phylogenetic relationships observed within the two species (Lewis-Rogers et al., 2009). All HRV-B and most HRV-A are members of the major receptor group. All HRV-B, except HRV-B27 are members of antiviral group A and all HRV-A, except HRV-A8 and –A95 are members of antiviral group B. Within HRV-A, the members of the minor receptor group form

several distinct clusters. However, these do not form a distinct genetic subgroup (Figure 1.2).

In 2006, a divergent phylogenetic clade of HRV variants were discovered from sequencing of HRV from respiratory samples (Lamson et al., 2006; Renwick et al., 2007; Arden et al., 2006; McErlean et al., 2007; Lau et al., 2007; Kistler et al., 2007a; Lee et al., 2007). These viruses were originally referred to as HRV-A2 or HRV-C and they received approved status by the ICTV as a distinct HRV species in 2009 (called HRV-C). HRV-C are genetically divergent from HRV-A and –B, except in the 5ˈUTR where there was evidence for inter-species recombination with HRV-A (Huang et al., 2009; Wisdom et al., 2009a). HRV-C have proven to be refractory to *in vitro* cell culture by classical methods used for isolation of HRV (McErlean et al., 2008). Recently, a cell culture system based on propagation of HRV-C in sinus mucosal organ culture has been developed and although this appears to be sensitive for HRV-C detection, the organ culture method is not practical for widespread use (Bochkov et al., 2011). Due to the lack of a suitable system for extensive analysis of the biological and replicative properties of HRV-C, these viruses have not been added to the classification systems described for HRV-A and HRV-B (including receptor usage, antigenicity and drug sensitivity profiles). Instead, classification of HRV-C must necessarily rely on nucleotide and amino acid sequence data derived from molecular studies.

FIGURE 1.2: Neighbour joining phylogenetic tree showing the prototype strains of HRV-A and corresponding biological properties. *Tree was constructed as described in Chapter 2. The majority of HRV-A strains are members of the major receptor group. Members of the minor receptor group are marked with a black dot. Members of antiviral group A are marked by a black box. Branches are scaled by genetic distance.*


HRV-A
Prototype strains
HRV-A15
HRV-A74
HRV-A38
HRV-A60
HRV-A67
HRV-A32
HRV-A9
HRV-A19
HRV-A82
HRV-A22
HRV-A64
HRV-A94
HRV-A61
HRV-A96
HRV-A73
HRV-A13
HRV-A41
HRV-A43
HRV-A75
HRV-A23
HRV-A30
HRV-A2
HRV-A49
HRV-A34
HRV-A50
HRV-A18
HRV-A57
HRV-A21
HRV-A55
HRV-A24
HRV-A90
HRV-A11
HRV-A33
HRV-A76
HRV-A16
HRV-A81
HRV-A1B
HRV-A1
HRV-A40
HRV-A85
HRV-A56
HRV-A54
HRV-A98
HRV-A59
HRV-A63
HRV-A39
HRV-A100
HRV-A10
HRV-A66
HRV-A77
HRV-A31
HRV-A47
HRV-A25
HRV-A62
HRV-A29
HRV-A44
HRV-A36
HRV-A89
HRV-A58
HRV-A7
HRV-A88
HRV-A53
HRV-A28
HRV-A68
HRV-A20
HRV-A71
HRV-A65
HRV-A51
HRV-A80
HRV-A46
HRV-A12
HRV-A78
HRV-A45
HRV-A95
HRV-A8
0.05

1.3 General aspects of picornavirus biology

1.3.1 Structure of picornavirus virions

Picornaviruses are very small RNA viruses of around 30nm in diameter. They are non-enveloped and therefore their infectivity is not affected by inorganic solvents. Despite the genetic heterogeneity observed with the capsid coding sequences of picornaviruses, the basic structure of the capsid is conserved throughout the family (Racaniello, 2007). The three dimensional structure of the virus capsid has been determined for such diverse picornaviruses as HRV (Figure 1.3), Seneca Valley Virus and FMDV (Oliveira et al., 1993; Garriga et al., 2012; Hadfield et al., 1997; Medappa et al., 1971; Venkataraman et al., 2008).

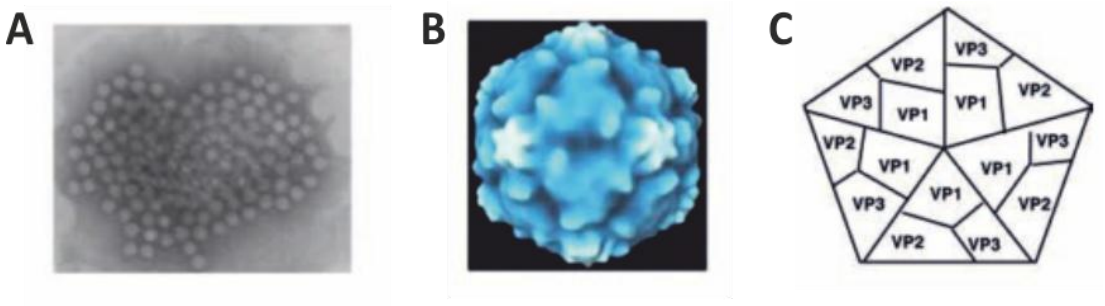


FIGURE 1.3: The structure of human rhinovirus virions. Adapted with permission from (Dreschers et al., 2007). A: Electron micrograph of purified rhinoviruses. B: Structural model of rhinovirus capsid obtained from analysis of cryoelectron microscopic data. C: Relative positions of capsid proteins VP1, VP2 and VP3 on the rhinovirus capsid.

The capsid is composed of 60 protomers, each containing one copy of the four capsid proteins VP1 to VP4. As one of the few exceptions, parechoviruses contain only three capsid proteins, as VP2 and VP4 are retained in their uncleaved precursor form VP0 (Stanway et al., 1994). The external capsid proteins, VP1 to VP3 form an 8 stranded antiparallel beta barrel (known as a "jelly roll" domain) which is a common

feature to other members of the order *Picornavirales* (Le Gall et al., 2008). The VP4 protein is located towards the internal aspect of the capsid. Groups of five protomers are arranged into 12 pentamers and these pentamers form a rigid pseudo-spherical structure with icosahedral symmetry. This configuration balances the stability and durability required for transmission with the flexibility required for cell surface binding and virion disassembly. Several studies have suggested that the picornavirus capsid may be a dynamic structure, which transiently exposes internal proteins (such as VP4 and the N-terminus of VP1) in a "breathing" process (Li et al., 1994; Lewis et al., 1998). This is important for release of the RNA genome. Capsid breathing also has implications for the development of antiviral therapies, as not only can antibodies raised against these internal regions neutralize infectivity in PV-1 but these can also confer cross-serotypic neutralization in HRV due to the highly conserved nature of this protein (Katpally et al., 2009; Li et al., 1994).

The virion has three axes of symmetry; a five fold, three fold and two fold axis. In most picornaviruses, a depression is formed near the five fold axis of symmetry by the beta barrels of the VP1 protein and this is known as the canyon. The canyon is the receptor binding site for major (but not minor) group HRV (Hewat et al., 2000; Colonno et al., 1988). Apthoviruses have been found to lack the canyon site and instead the receptor binds to a flexible loop which projects from the surface of the virion (Acharya et al., 1989). In EV, immediately below the canyon floor lies a hydrophobic pocket region which contains a "pocket factor" (Oliveira et al., 1993; Verdaguer et al., 2000). This region is a drug target for various antiviral drugs and successful binding has been shown to inhibit the uncoating of the RNA genome and subsequent delivery into the host cell (Smith et al., 1986).

1.3.2 Genome organisation and proteolytic processing

The genome of all picornaviruses consists of single-stranded, positive sense, non-segmented RNA of around 7 to 9 kilobases (kB). The genome comprises of a single

coding region which is flanked by 5' and 3' UTRs of varying lengths. In addition, the 3' polyA tail also varies significantly in length between genera and has been shown to be required for infectivity (Spector, 1974; Racaniello, 2007). The VPg (viral protein genome-linked) is a small protein that is covalently linked to the 5' end of the genome and acts as a primer for genome replication. The VPg is encoded by the short and well conserved 3B genome region (Figure 1.4). In contrast to other picornaviruses, FMDV and aquamavirus A encode more than one VPg molecule (Forss and Schaller, 1982; Knowles and Wadsworth, 2010).

The picornavirus genome consists of a single open reading frame translated as a single polyprotein. However, the complete polyprotein is cleaved co-translationally and so is never evident in the cell (Racaniello, 2007). Unlike mammalian RNA, picornavirus RNA lacks a methylated 5' cap structure and initiation of translation at the ribosome is carried out by the IRES within the 5'UTR. The translated polyprotein is divided into capsid and non-structural regions, known as P1, P2 and P3 (Rueckert and Wimmer, 1984). These regions encode four structural and seven non-structural proteins, which are cleaved by virus-encoded proteinases in a cascade of proteolytic processing events (Figure 1.4). The cascade shown is specific to the sequence of events in EV in keeping with the focus of this thesis. However, several notable differences between proteolytic processing events in other picornavirus genera exist.

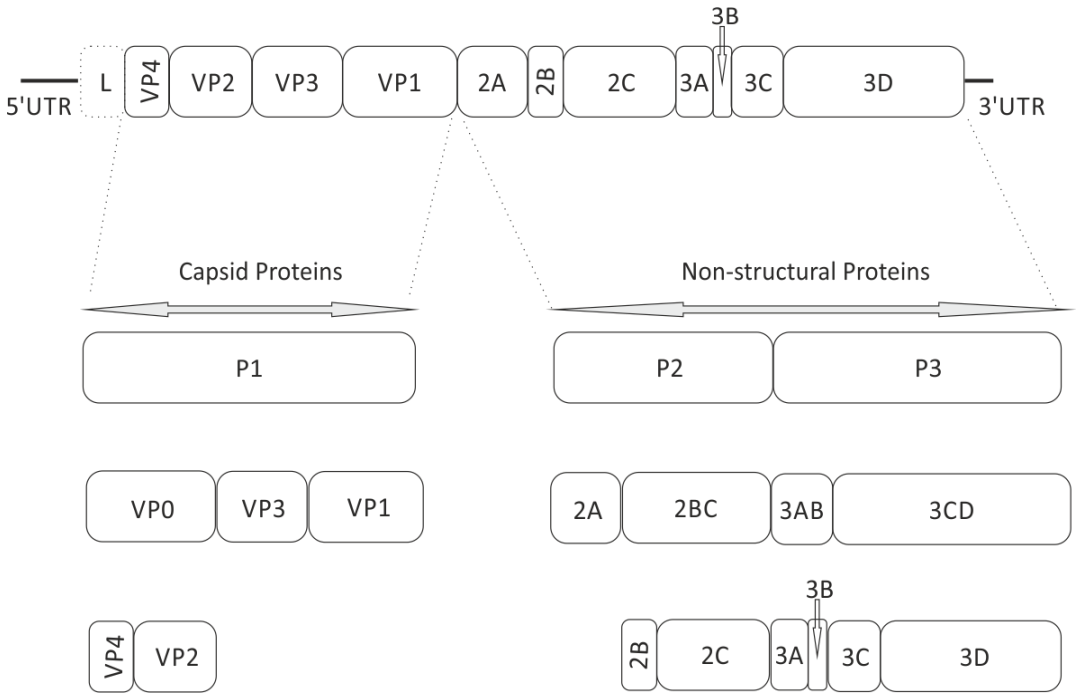


FIGURE 1.4: Sequence of cleavages during proteolytic processing of the translated genome of enteroviruses. *Organisation of picornavirus genome is shown on the top line. Relative sizes of protein products are taken from annotation from Genbank entry for HRV-B14 : X01087. The Leader protein (L) is not present in EV and position in the genome is shown for reference only. Intermediate and final protein products are shown.*

The junction between the P1 and P2 coding regions is cleaved in EV by the 2A proteinase. However, in certain other genera, including *Cardiovirus*, *Parechovirus*, *Hepatovirus* and *Apthovirus*, this junction is cleaved by the 3C proteinase (Racaniello, 2007), as the 2A protein does not have proteolytic activity. Several picornavirus genera (*Apthovirus, Cardiovirus, Erbovirus, Kobuvirus* and *Sapelovirus)* encode an additional leader protein (L), at a 5' position to the VP4 capsid protein (position shown by dotted box in Figure 1.4). This functions as a proteinase in apthoviruses and has a role in the inhibition of host-cell protein synthesis (Devaney et al., 1988). However, in cardioviruses, the leader protein lacks proteolytic activity and is instead cleaved from the polyprotein by 3C. Other cleavage events, including the cleavage of P2 and P3, are undertaken either by the 3C proteinase or $3CD^{pro}$, an intermediate protein product that functions as a

proteinase. The encoded proteins 2B, 2BC and 3AB are all membrane associated proteins with roles in viral replication (VR Racaniello, 2007). VP1 to VP4 assemble into the final capsid structure and the last event in the assembly of the picornavirus virion is the cleavage of VP0 to form VP4 and VP2.

1.3.3 RNA secondary structure in picornavirus genomes

In addition to the encoded proteins, picornavirus genomes contain several well defined areas of functional RNA secondary structure. RNA secondary structure is formed by base pairing of single stranded RNA with itself and leads to the formation of structures such as stem-loops and pseudoknots. RNA secondary structure can form randomly or can have a high degree of functional importance and in practice it is often challenging to separate these two.

Certain structural elements are present in picornavirus genomes, including elements within the UTRs and the *cis-acting* replication element (cre) within the coding region. The cre is an RNA secondary structure which has been observed within the genome of several picornavirus genera. It has been implicated in negative strand RNA synthesis (Goodfellow et al., 2000) and in VPg uridylation, allowing VPg to function as a primer for genome replication (Paul et al., 2000). The location of the cre has been pinpointed in several different picornavirus genera and there is a high degree of structural conservation. Despite this, cre elements from different species exhibit high nucleotide sequence divergence and are located in different parts of the coding or non-coding regions of the genome (Cordey et al., 2008; Goodfellow et al., 2000; Witwer et al., 2001; McKnight and Lemon, 1998; Gerber et al., 2001). In fact, it has been suggested that this variability in position may be used as an additional classification tool for identifying species of HRV and EV (Cordey et al., 2008).

The IRES is a long and highly structured stretch of nucleotides within the 5'UTR. Picornavirus genomes lack the 5' cap structure observed in mammalian RNA, which is integral to ribosome binding. The IRES was first described as the site of internal

initiation of translation for picornavirus genomes in 1988 (Pelletier and Sonenberg, 1988; Jang et al., 1988). There are five classes of viral IRES and two main types are present in picornaviruses. The type I and type II IRES of EV and apthoviruses/cardioviruses respectively have little homology in either structure or sequence and yet provide the same basic function (Racaniello, 2007) (Figure 1.5). Hepatitis A virus (HAV) contains a further distinct IRES, designated type III (Belsham, 2009). In addition, several other picornavirus genera (including *Avihepatovirus, Teschovirus, Senecavirus, Sapelovirus* and *Tremovirus*) have been shown to contain IRES elements which are similar to those observed in hepatitis C virus and pestiviruses in the virus family *Flaviviridae* (Willcocks et al., 2011; Hellen and de Breyne, 2007). Although it is possible that this IRES type (type IV; Figure 1.X) may have evolved independently, it has also been suggested that this may have arisen by recombination and modular exchange of replication elements between members of different virus families (Hellen and de Breyne, 2007).

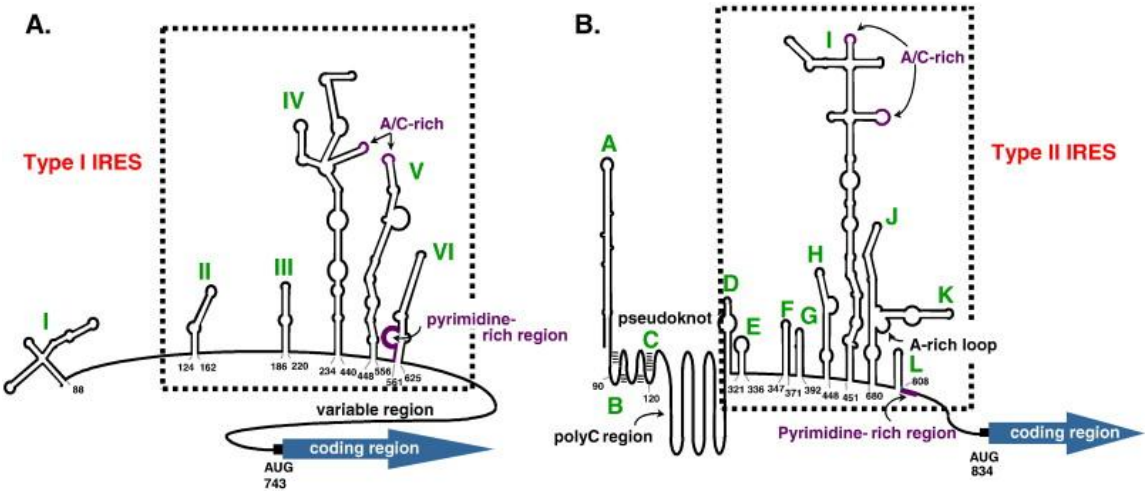


FIGURE 1.5: RNA secondary structure of picornavirus IRES elements. Presented with permission from (Fitzgerald and Semler, 2009). A: Type I IRES, as observed in *Enterovirus*. B: Type II IRES, as observed in *Apthovirus*.

In addition to these two regions of RNA secondary structure present in all picornaviruses, the 3'UTR contains elements of secondary structure thought to be

required in viral replication (Brown et al., 2005) and a 5' cloverleaf structure has been observed in the 5'UTR of enteroviruses (Palmenberg et al., 2009). This 5' structure has been implicated in the initiation of RNA synthesis (Barton et al., 2001). The entire 5'UTR of EVs can be divided into six functional domains, consisting of the cloverleaf structure (domain I) and the IRES (domains II – VI) (Rohll et al., 1994). Within the cloverleaf, four individual stem-loops exist (A-D). Stem loop B specifically binds with poly(rC)-binding proteins and stem loop D interacts with the viral 3C and 3CD proteases (Du et al., 2004). These have been implicated in the switch between translation and replication (Rohll et al., 1994).

1.3.4 Life cycles of picornaviruses

The replication of picornaviruses takes place in the cytoplasm of the host cell. Attachment of the virus to a specific cell surface receptor leads to structural changes in the capsid which allow the uncoating of the RNA genome and its subsequent entry into the cell. The genome is first translated to provide copies of the viral proteins required for genome replication. Replication is preceded by negative strand synthesis and this strand is then used as a template for positive strand RNA synthesis. Capsid proteins and newly synthesized viral genomes assemble into mature virions and are then released from the cell. The entire cycle is complete in around 5 to 10 hours (Racaniello, 2007) (Figure 1.6).

Attachment of picornaviruses to host cells occurs via a wide variety of cell surface receptors. Cell surface receptor specificity is an essential determining factor in cell/tissue tropism and therefore both disease manifestations and host species. However, many cellular receptors exploited by picornaviruses are expressed in a wide variety of tissues and therefore tropism may not be primarily determined by receptor expression alone. For example, both I-CAM1 and CD155, members of the immunoglobulin superfamily, are expressed in many tissue types throughout the

body and yet, the primary sites of replication of HRV and PV are the respiratory and gastrointestinal tracts respectively.

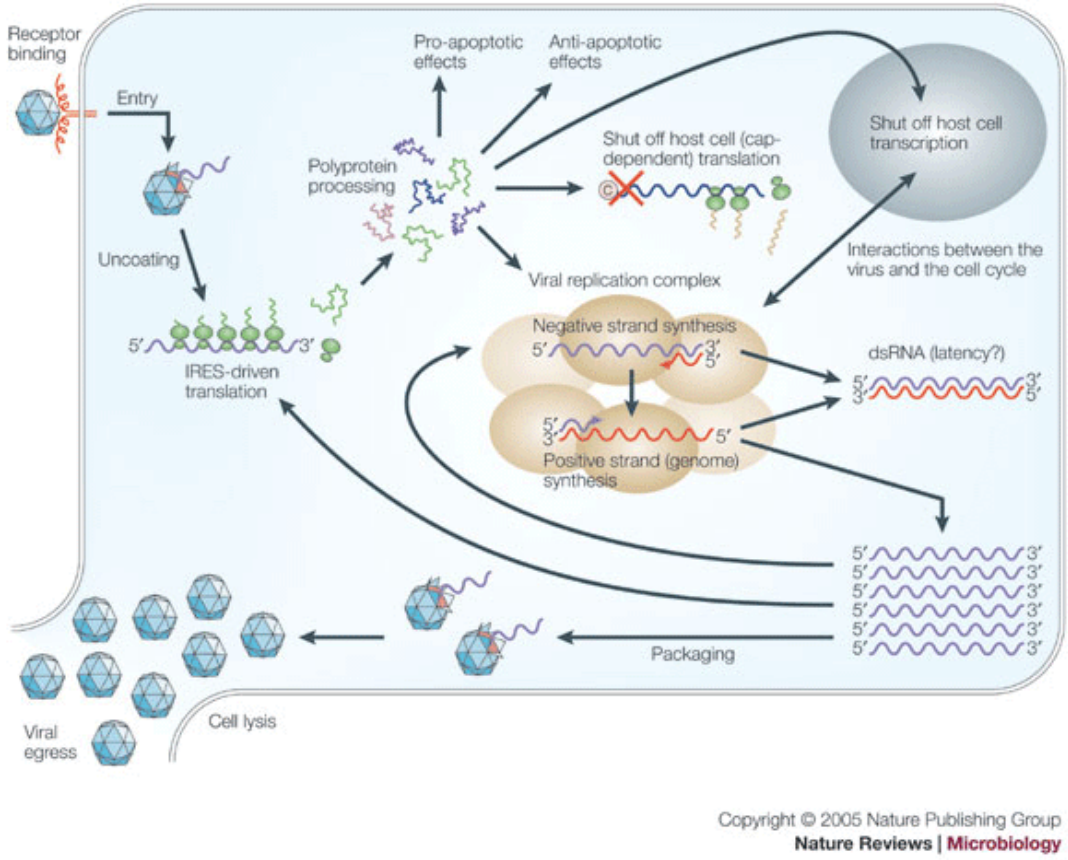


FIGURE 1.6: Summary of the picornavirus life cycle. Presented with permission from (Whitton et al., 2005). The figure summarises key events in the replication of picornaviruses and the main events are described in the text below. © represents the cap present on most cellular mRNA.

The receptor usage profile of picornaviruses is complex and receptor specificities of some major picornavirus species are summarised in Table 1.2. Some EV and HRV utilise receptors which are members of the immunoglobulin superfamily and binding occurs in the canyon region (Colonno et al., 1988). This leads directly to the conformational changes in the capsid required for uncoating of the genome.

TABLE 1.2: Cell surface receptors used by selected picornaviruses

Receptor family	**Receptor Type**	**Virus**	**Natural host**	**Primary site of replication**
Immunoglobulin superfamily	I-CAM1	HRV	Human	Respiratory tract
	CD155	PV	Human	Gastrointestinal tract
	Coxsackie and adenovirus receptor (CAR)	Coxsackie viruses	Human	Gastroinestinal tract
	Murine vascular cell adhesion molecular 1 (VCAM-1)	EMCV	Pig	Gastrointestinal tract
Decay Accelerating Factor (DAF)	DAF/CD55	Some enteroviruses	Human	Gastrointestinal tract
Low density lipoprotein family	LDL	HRV	Human	Respiratory tract
Integrins	Integrins	FMDV Some enteroviruses	Cloven footed animals Humans	Respiratory tract Gastrointestinal tract
Sialic acid	Sialic acid	Cardioviruses ERAV	Rodents Horses	 Respiratory tract
EV71 receptors	Human P-selectin glycoprotein ligand 1	EV71	Humans	Gastrointestinal tract
Hepatitis A virus receptor	HAV-cr1	HAV	Humans	Gastrointestinal tract

- Information adapted from (Tuthill et al., 2010; Racaniello, 2007; Pallansch and Roos, 2007)

Binding of the cell surface receptor leads to uncoating of the viral genome. Unlike enveloped viruses which can directly fuse their lipid coat with the host cell plasma membrane, the non-enveloped picornaviruses must exploit a variety of different mechanisms in order to effectively deliver their genome into the cytoplasm. Picornaviruses do not enter the cell directly through the plasma membrane, instead relying on endocytosis and subsequent release from vesicles (Brandenburg et al., 2007; Tuthill et al., 2010). The mechanism by which endocytosis takes place and the cue utilised for timely uncoating of the RNA varies between viruses. In all cases, the

plasticity of the conformation of the capsid and the phenomenon of "capsid breathing" is of crucial importance to this process (Roy and Post, 2012).

Endocytosis can be clathrin-mediated, as is observed for apthoviruses, some HRV and HAV (Tuthill et al., 2010). Some EV undergo endocytotic uptake via a caveolin mediated pathway, whereas PV depends on a non-caveolin non-clathrin mediated process (Brandenburg et al., 2007; DeTulleo and Kirchhausen, 1998). The actual uncoating of the RNA genome depends on either a direct mechanical cue secondary to receptor binding (as in major group HRV) or a chemical cue such as low pH (Suomalainen and Greber, 2013). This occurs at around pH 5.4 in the minor group HRV (Brabec et al., 2003). The N-myristoylated VP4 capsid protein and the N-terminus of the VP1 protein, usually present on the internal surface of the virus, are externalised. The exposed segment of the VP1 protein tethers the virus to the membrane, while the VP4 protein can form a pore which is suggested to allow the release of the genome into the cytoplasm (Davis et al., 2008; Fricks and Hogle, 1990; Danthi et al., 2003). Post release of the genome, the 80S empty particle of EV is still present within the endosome. However, the capsids of apthoviruses and cardioviruses dissociate into respective pentameric subunits (Tuthill et al., 2010). Uncoating of picornaviruses with a canyon region within the capsid can be blocked by certain antiviral compounds, such as pleconaril or WIN compounds. These bind to the canyon and stabilize the capsid conformation, thereby blocking the release of VP4 (Roy and Post, 2012).

Once in the cytoplasm, the viral RNA must be first translated in order to provide copies of the viral proteins necessary for genome replication, which are not naturally present in the host cell. Once sufficient viral proteins have been synthesized, the switch from translation to replication occurs, as both cannot occur on the same RNA template simultaneously (Racaniello, 2007). The viral RNA dependent RNA polymerase, encoded by the 3D genome region, is responsible for RNA synthesis.

Both the cellular protein poly(rC) binding protein and the viral 3CD protease bind to the 5' cloverleaf structure to form the ribonucleoprotein complex (Rohll et al., 1994; Andino et al., 1990). A protein-protein bridge is then formed between this complex and the cellular polyA binding protein that is bound to the 3' end of the viral genome (Herold and Andino, 2001). This has the effect of circularizing the RNA, which in turn initiates negative strand RNA synthesis from the 3'polyA region. Negative strand RNA synthesis results in the formation of a positive/negative RNA duplex known as the replication form. Positive strand synthesis then occurs using the negative strand as a template and secondary to the anchoring of the negative strand by virally encoded protein 2C. Uridylated VPg is required as a primer for RNA replication. RNA replication takes place within replication complexes bound to smooth membrane vesicles (Bienz et al., 1994). The compartmentalization of RNA replication may increase efficiency of replication by localising all required components and additionally may facilitate the avoidance of eliciting a dsRNA activated immune response during the replication cycle (Racaniello, 2007).

Some of the newly produced positive stranded RNA viral genomes are translated to produce viral capsid proteins. Each positive stranded RNA genome is linked to a VPg protein and encapsidated. The final event before cell lysis is the maturation cleavage of capsid protein VP0 to form VP4 and VP2 (Lee et al., 1993).

1.3.5 Pathogenesis of enterovirus infections

Members of the genus *Enterovirus* comprise a large number of pathogens of both humans and animals (Table 1.1). Although the human EV have been documented to cause such varied clinical presentations as respiratory infections, hepatitis, acute haemorrhagic conjunctivitis and myocarditis (Gaaloul et al., 2012; Pallansch and Roos, 2007), the main public health concern for this group of viruses is their propensity to cause infections of the central nervous system (CNS). These include the well-known poliomyelitis (caused by PV), a large number of different paralysis

syndromes, aseptic meningitis and encephalitis (Siafakas et al., 2001; Kupila et al., 2006). Members of all four species of human EV have been documented to cause infections of the human CNS.

The CNS is protected from pathogens and toxins in circulating blood by the blood-brain-barrier. This consists largely of endothelial cells which regulate the transport of molecules in and out of the cerebrospinal fluid (CSF). There are two postulated mechanisms of EV entry into the CNS; either through haematogenous spread and subsequent violation of the blood brain barrier or invasion of peripheral nerves leading to retrograde axonal flow to the brain (Ren and Racaniello, 1992; Ohka et al., 1998). The phenomenon of retrograde axonal flow, whereby virus gains access to the CNS via invasion of peripheral nerves has been specifically observed for PV infections. In general, invasion of the CNS by EV is thought to occur due to haematogenous spread. Upon entry via the faecal-oral route, EV initially replicates within the associated lympatic tissues of the small bowel and subsequent invasion of the blood vessels leads to viraemia. Viraemia can occur even in cases of asymptomatic EV infection (Welch et al., 2001). Excretion of the virus occurs in faeces for weeks after the initial infection and occurs in both symptomatic and asymptomatic infections.

Human EV notably possess acid stable virions and are therefore able to pass through the highly acidic environment of the stomach prior to establishing infection in the small bowel. In contrast, HRV are thought to be acid labile (Giranda et al., 1992) and therefore have a respiratory tropism. The same association is observed in apthoviruses, which are acid labile viruses that replicate predominantly in the respiratory tract. HRV also appear to propagate *in vitro* preferentially at 33°C, which has been interpreted as an adaptation to the cooler environment of the upper respiratory tract. However, recent studies have indicated that HRV are capable of replication both at the higher temperatures associated with the lower respiratory tract

and within the lower respiratory tract itself (Gern et al., 1997; Mosser et al., 2002; Papadopoulos et al., 2000).

1.4 Evolution of RNA viruses

1.4.1 General points in the evolution of RNA viruses

The genetic diversification and evolution of viruses occurs by three main mechanisms; point mutations, recombination and reassortment. Point mutations cause small changes in the genome sequence during replication, secondary to the error prone nature and lack of proof-reading activity of the viral polymerase (Steinhauer et al., 1992). Recombination and reassortment cause the creation of chimaeric genomes, by transferring large genome regions between different viruses. Reassortment involves the transfer of entire genome segments and therefore occurs only in viruses with segmented genomes (such as influenza). Reassortment is not discussed in the brief review below.

RNA viruses generally have short replication cycles which allow the production of many virions, including a great number of mutants in each cycle. The fixation of mutations in these viruses leads to divergent RNA sequences and therefore frequently altered protein sequences. Although a large number of mutants are generated with each replication cycle, all are required to conserve essential functions to maintain viability. All mutant genomes are subjected to the same selective pressures and the observed genetic diversity of a viral population depends not only on the mechanism of generation of mutants, but on their fitness and ability to produce infectious progeny. Some RNA viruses are thought to exist as a swarm of closely related mutants clustering around a general consensus sequence, known as a quasispecies (Domingo et al., 1996, 2008). The rapid generation of diversity means that viral populations can quickly adapt to changing environments, which has implications for the development of both antiviral therapies and vaccines.

Constraints on observed genetic diversity are imposed in several ways. Not only does every mutation or recombination event have implications for viral fitness, but the total genome size of picornaviruses is constrained by the fixed internal space within the viral capsid. However, PV with up to an additional 1500 nucleotides have been shown to be successfully packaged (Alexander et al., 1994). In addition, viral evolution can have important consequences for virulence and a balance must be struck between successful viral propagation and causing the premature death of the host cell.

1.4.2 Point mutations

Nucleotide misincorporation is estimated to occur at a frequency as high as one mutation per genome per round of replication in RNA viruses (Drake and Holland, 1999). The existence of picornaviruses as a related swarm of micro-variants or quasispecies means that there is a large pool of slightly mutated genomes present at any one time, which may become fitter variants as the selection pressures exerted by the host change. Picornaviruses are a highly genetically diverse group of viruses, both within and between species. Despite the fact that most random mutations are likely to be deleterious, it has been found that increasing the fidelity of the RNA polymerase actually causes attenuation of PV (Pfeiffer and Kirkegaard, 2005). However, decreasing the fidelity of the polymerase will necessarily lead to the creation of a large number of non-viable progeny and eventual complete inability to replicate (error catastrophe). For example, the small molecule ribavirin has been demonstrated to cause error catastrophe in PV by reducing the fidelity of the polymerase (Crotty et al., 2001). Therefore, random mutations in RNA viruses are a delicate balance between generating sufficient viral diversity to adapt to a rapidly changing host environment and avoidance of crossing the error threshold.

Point mutations occur by the random introduction of either single or multiple nucleotide changes during RNA replication. Even single base mutations can have

profound effects on viral phenotype and therefore these small, incremental changes are subject to the evolutionary forces of positive and negative selection. A progeny genome with a lowered fitness will be outcompeted by other variants in the population and therefore is likely to disappear quickly. Random non-synonymous mutations (which cause a change in amino acid sequence) can decrease viral fitness by altering the essential secondary structure or interactions of the encoded proteins. Insertions and deletions in the nucleotide sequence (indels) are significantly less commonly observed in sequenced viral genomes than single base point mutations. This is due to the fact that the addition or removal of a single base will cause a frameshift mutation resulting in a protein that is non-functional or truncated (if the indel introduces a stop codon). In order to maintain viability of the virus, the indel must occur in a region of genome that is not functionally conserved and must result in the insertion or deletion in multiples of three to preserve downstream coding.

Different segments of the genome have distinct roles in the viral life cycle and therefore are subjected to differing evolutionary pressures. For example, a staggeringly high genetic diversity is observed within the majority of the capsid coding region of both HRV and EV (McIntyre et al., 2010; Lindberg et al., 2003; Laine et al., 2006). This is likely secondary to the fact that these proteins are exposed to the host immune system and the generation of diversity, especially in the binding sites of neutralizing antigens likely allows immune escape.

1.4.3 Recombination

Genetic diversity is also routinely created in RNA viruses by recombination. Recombination is the joining of two genome regions from related or unrelated viruses to create a chimaeric genome. This results in both new viral phenotypes which may obtain an evolutionary advantage and also is thought to decrease the load of deleterious mutations (Worobey and Holmes, 1999).

Recombination can be inferred indirectly from analysis of sequence data, as different recombinant genome regions will display different evolutionary histories. Although computational and phylogenetic methods in use today for the detection of recombination are both powerful and accurate, recombination between very closely related parents cannot be detected. Non-viable recombinants that are likely generated randomly within a single replication cycle also cannot usually be detected. Recombination within RNA viruses can be classified as homologous, aberrant homologous or nonhomologous (Lai, 1992). Homologous recombination occurs within a homologous region of two related viruses. Due to the likelihood of these generated recombinants being viable and able to propagate, it is these that are most often observed. Aberrant homologous recombination occurs between homologous viruses that are not in strict alignment and the relatively infrequent nonhomologous recombination occurs between two completely unrelated viruses.

The most widely accepted model of recombination in RNA viruses is known as the copy-choice model, whereby the RNA dependent RNA polymerase switches template strand during synthesis of the negative strand (Kirkegaard and Baltimore, 1986). This mechanism is necessarily replication dependent and relies both on co-infection of the same host cell by two different viruses and replication of the two viruses within the same cellular compartment. The dissociation and resulting switching may be related to "pausing" of the polymerase secondary to regions of RNA secondary structure or nucleotide misincorporations (Agol, 1997). A second mechanism is proposed whereby recombination occurs by the breaking and re-joining of RNA strands (Gmyl et al., 2003) and this is not replication dependent.

Recombination has been extensively documented in a variety of picornavirus genera, including *Apthovirus* (Heath et al., 2006), *Parechovirus* (Benschop et al., 2008; Calvert et al., 2010), *Teschovirus* (Simmonds, 2006) and *Enterovirus* (Oprisan et al., 2002; Santti et al., 1999; Simmonds and Welch, 2006). Within these picornavirus

genera, documented recombination events have tended to occur in the 5ˈUTR and the non-structural coding regions. Natural recombination of vaccine PV strains with circulating EV is also widely responsible for the generation of vaccine escape mutants and subsequent outbreaks of vaccine associated poliomyelitis, which has grave implications for the PV eradication campaign (Jegouic et al., 2009; Cuervo et al., 2001; Guillot et al., 2000). The oral live-attenuated poliovirus vaccine (Sabin) contains three serotypes of poliovirus (PV-1, -2 and -3). After administration of the oral vaccine, replication occurs in the gut tissues and thereby elicits a strong, local immune response preventing subsequent infection with wild poliovirus via the faecal-oral route. Unfortunately, the nature of the live-attenuated vaccine presents significant opportunities for inter-typic recombination of PV strains during replication. Recombinant PV strains have been widely linked to cases of vaccine associated paralytic poliomyelitis and documented recombination partners have included other Sabin vaccine strains (Cuervo et al., 2001), circulating wild poliovirus (Dahourou et al., 2002) and other related EV-C types (Jegouic et al., 2009). Understanding the frequency and nature of recombination events occurring in the natural circulation of a virus is vital for potential development and implementation of therapeutic strategies and potential vaccines.

1.5 Aims of this thesis

With the discovery of the novel HRV-C species in 2006 and the gradual realisation that HRV are responsible for a much wider variety of more severe clinical illness than was previously thought has come a renewed interest in research of the clinical and biological features of HRV. This thesis comprises of several distinct studies of the epidemiology, classification and evolution of HRV.

One of the primary goals of this research was to conduct preliminary explorations of other potential clinical manifestations of HRV infections. Although HRV are

traditionally thought to be acid labile and are therefore to only infect the respiratory tract, acid-stable mutants have been previously isolated (Skern et al., 1991). The closely related human EV are an established and important cause of aseptic (viral) meningitis and are frequently isolated from stool specimens. In light of this fact, CSF and stool samples were screened for HRV, in order to determine whether the virus was present and if so, whether any epidemiological patterns could be elicited. In addition, concurrent studies were undertaken of circulating HRV epidemiology in respiratory samples isolated from Edinburgh. These studies are outlined in Chapter 3 and aspects pertaining to HRV isolation from stool and CSF were published as part of a larger study in 2012 (Harvala et al., 2012b).

The numerous studies of HRV epidemiology undertaken across the globe has led to the rapid accumulation of a vast amount of HRV sequence data. Indeed, in the analysis of sequence data acquired from our own epidemiological studies and published sequences, a large number of sequence groups were identified that appeared to be divergent from known HRV types. At the time that this work was initially undertaken, the main method of classification of HRV and assignment of putative new types was by studies of neutralisation properties. These methods were not suitable for the study of HRV-C, as the virus cannot be isolated in standard cell lines and at that time, no alternative means of *in vitro* propagation existed. However, a system had been in use for several years, whereby EV could be classified and novel types assigned by nucleotide sequence divergence in the VP1 region (Oberste et al., 1999a, 1999b). Chapter 4 focusses on the determination of nucleotide sequence divergence thresholds in two capsid coding regions of HRV and their subsequent application in both defining new types and redefining previous erroneous classifications. Our proposals for the classification of HRV-C into genotypically defined types was published in 2010 and has now been adopted by the Picornavirus Study Group for the classification of this virus species (Simmonds et al., 2010). In

addition, we have formulated proposals for the extension of this classification system to include HRV-A and HRV-B (McIntyre et al., 2013a).

The classification of all available HRV sequence data into types defined by phylogenetic relationships and sequence identity allowed the completion of studies of the evolution of HRV species and types. Although the phenomenon of recombination is frequently observed in EV, several studies had suggested that it was relatively rare in the evolution of HRV coding genomes (Kistler et al., 2007b; Lewis-Rogers et al., 2009). Studies of the occurrence of recombination within the coding region of HRV species which included many isolates of the same type had not, to our knowledge, been undertaken. However, prior evidence did exist for widespread recombination between the 5ˈUTR regions of HRV-A and HRV-C (Huang et al., 2009; Wisdom et al., 2009a). Chapter 5 describes an analysis of recombination within the coding region and genetic diversity of HRV-C strains and includes an in-depth analysis of recombination in the 5ˈUTR. This analysis was published in 2010 (McIntyre et al., 2010).

However, the vast majority of HRV-C sequences generated for this study originated from the same 12 month period within the same geographical location. It has been previously noted that even the frequently recombinant EV-B sequences are not always evident as such during a short time-scale (McWilliam Leitch et al., 2009a). Therefore, the analysis of recombination in HRV was extended to include samples from all three HRV species and covering a time-span of at least 14 years. The study of the occurrence of recombination in the evolution of all three species of HRV is reported in Chapter 6 and was published in 2013 (McIntyre et al., 2013b).

Chapter 2
Materials and Methods

Materials and methods that were used in the completion of the studies within this thesis comprise two main strands. Firstly, laboratory techniques that were applied to both screening large number of samples for the presence of HRV and producing sequence fragments from known positive isolates for evolutionary studies are described in Section 2.2. The second strand comprised computational analyses directed towards the construction of comprehensive datasets, phylogenetic and recombination analyses and determination of mean substitution rates in specific genome regions. These are described in Section 2.3. Also central to both areas of study were a series of archives of clinical specimens at the Scottish Virology Centre (SVC), described in Section 2.1.

2.1 Storage and archiving of clinical samples

Since 2005, large archives of clinical samples referred for virological testing have been maintained in collaboration between the SVC based at the Royal Infirmary of Edinburgh (RIE) and the University of Edinburgh. Referral of patient samples for diagnostic virology is generally undertaken for system-specific suspected diagnoses (for example, pneumonia for respiratory samples and aseptic meningitis for CSF samples). Although certain situations exist where the referral of samples is strongly indicated, such as pyrexia of unknown origin or severe unexplained headache, the process is largely dependent on the clinical judgement of the attending medical staff. Although effective interventions do not yet exist for a large number of viruses that are routinely screened for, viral diagnostics are also essential for prediction of prognosis (including possible complications) in individual patients, creating infection control protocols and for disease surveillance (Templeton, 2007).

The collection and storage of these referred samples additionally facilitates the completion of many studies of viral epidemiology, clinical characteristics and evolution that would not otherwise be easily possible. Samples stored in each archive are allocated an anonymised number and any information which would render the individual patient identifiable are removed. This includes name, date of birth and Community Health Index (CHI)/patient number. Approval for the retention of specific, anonymised patient data for the purposes of epidemiological research was initially requested by Professor P. Simmonds in 2005 and granted by the Lothian Regional Ethics Committee (Development of molecular methods to enhance diagnostic testing for viruses and other infectious agents; REC reference number 08/S1102/2). Non-identifiable patient information is retained in the form of a secure (password protected) database. This includes age band of the patient, year and month of collection, patient location (whether based in an Intensive Therapy Unit (ITU), general ward or community health centre), sample type, basic clinical information and the results of routine diagnostic screening. As the basic clinical information recorded largely consists of details given on referral forms, data is frequently missing or incomplete. Prospective data collection for specific studies has allowed certain records to retain more detailed non-identifiable clinical information for specific cases. In addition, the results of other microbiology screening are recorded (Templeton et al., 2004; Scheltinga et al., 2005; Gaunt et al., 2010).

In parallel with the maintenance of the archive of original clinical samples, nucleic acid extracts produced during routine viral diagnostic procedures are stored in a related and cross-referenced archive. Original clinical samples are stored at -80°C, while extracted RNA/DNA is stored at either -45°C or -20°C. A supplementary archive of any complementary DNA (cDNA) produced is kept at -20°C. However, these cDNA samples are kept for further use only when they are made by amplification with random primers and therefore are useful for screening for other RNA viruses. With the exception of samples obtained from Finland (Chapter 6), all

clinical samples and HRV positive isolates used within the course of these studies were obtained from these clinical archives.

2.2 Laboratory techniques

2.2.1 Prevention of contamination of polymerase chain reaction products

The studies contained within this thesis relied on the generation of large amounts sequence data by polymerase chain reaction (PCR) amplification of genome fragments from HRV positive samples and subsequent Sanger sequencing. The exquisitely high sensitivity of PCR based techniques is potentially subject to contamination by extraneous DNA sequences and avoidance of this has been especially important for the work described herein. The accuracy of studies which include both screening large numbers of samples for a specific virus and amplifying non-consecutive genome fragments for evolutionary analysis can be severely compromised by PCR contamination.

Considerable precautions were taken in the completion of laboratory based work. In addition to employing standard good laboratory practice and practicing aseptic technique at all times, reagents and primers for the preparation of PCR mixes were kept in small aliquots and reserved for sole use of specific individuals where possible. In an attempt to prevent false positives resulting from contamination of adjacent PCR tubes in the preparation of reaction mixtures, DNA was added to the reaction mixture after all other reagents (Kwok and Higuchi, 1989). In order to detect potential reagent contamination, a negative control prepared from the same reagent mix as all actual samples was included with all PCR based experiments.

As even one round of PCR generates extremely high copy numbers of DNA fragments, one potential method of significantly reducing contamination is to limit carryover. Carryover occurs when PCR reaction mixtures are contaminated with

products of previous reactions. In order to limit this, we used separate areas and dedicated equipment for the various stages in the amplification of DNA from clinical samples (Figure 2.1). This comprised of a "one-way" laboratory system, where only unidirectional transfer of reagents, equipment or samples between certain areas was permitted (marked by filled arrows on Figure 2.1).

Sample preparation, including defrosting, processing, cataloguing and nucleic acid extraction was carried out in a dedicated area reserved specifically for this purpose. PCR buffers and reagents were prepared in a clean room, with separate laboratory and personal protective equipment. In the latter part of this work, positive controls comprised of known copy number RNA transcripts (McLeish et al., 2012). Due to the potential for contamination of PCR reactions with high copy number transcripts, serial dilutions of RNA transcripts were additionally carried out in a dedicated laboratory area, separated from the area reserved for preparation of PCR reagents.

cDNA synthesis and first round PCR reactions were carried out in the same area (marked with white arrows on Figure 2.1). In the majority of this work, second round PCR was carried out in a dedicated laboratory area, separate from gel electrophoresis and preparation of sequencing reactions. However, in later work, due to laboratory layout and space restrictions, backwards travel was necessary between the dedicated areas for these three procedures (marked by grey box on Figure 2.1).

The adherence of all group members to these protocols has ensured that the occurrence of PCR contamination observed within the completion of the studies contained within this thesis has been minimal.

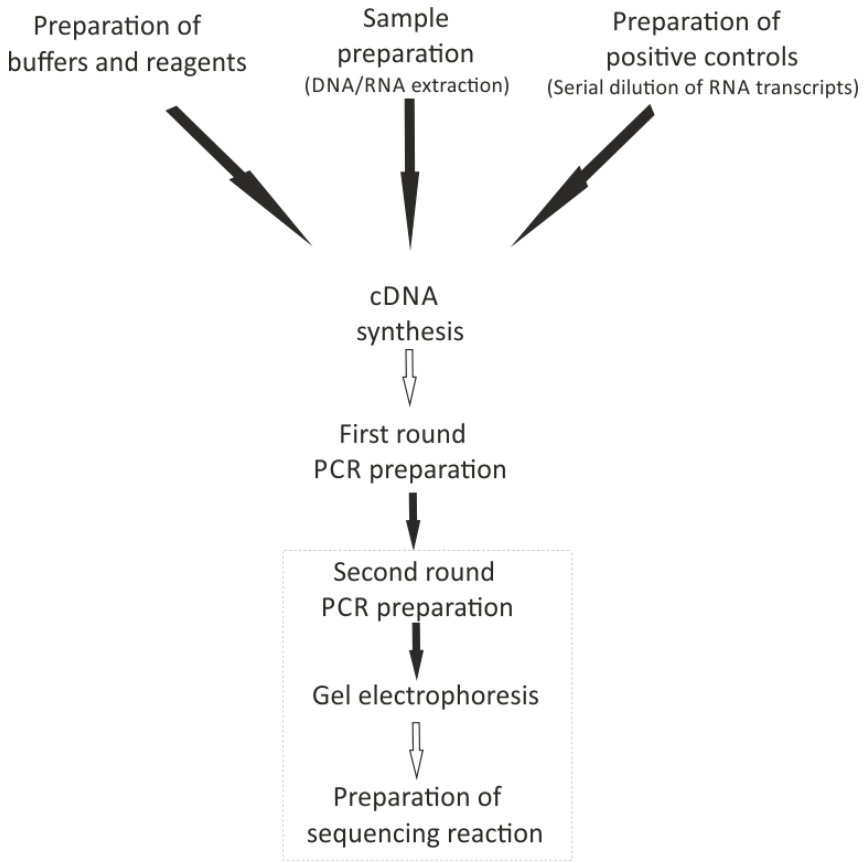


FIGURE 2.1: "One way" laboratory system for the prevention of carryover contamination in PCR reactions. *Filled arrows denote steps between areas which originally were designed to flow in one direction. Backwards travel of reagents, samples or equipment was not permitted between these areas. White arrows denote procedures which can be carried out in the same work-space. In later laboratory work, due to space restrictions, second round PCR preparation and preparation of sequencing reactions were required to be carried out within the same laboratory area (marked with grey box). An effort was still made to separate physical bench space, reagents and pipettes for these procedures.*

2.2.2 Extraction of RNA from clinical samples

In cases where nucleic acids extracted during routine diagnostics at the SVC were not available for use in research studies, RNA was directly extracted from clinical specimens using the Qiagen Viral RNA Mini Kit (Qiagen, UK), according to the manufacturers' instructions. Samples were stored at -40°C and aliquots were defrosted in a microbiological safety cabinet. Stool samples were thoroughly mixed with phosphate buffered saline (PBS) and clarified by low-speed centrifugation. The

resulting supernatant was used in extraction procedures. Respiratory and CSF samples were used directly in the extraction protocol.

An identical protocol for RNA extraction was used in the processing of all sample types. Briefly, 140µL of sample was incubated at room temperature with 560µL viral lysis buffer (Buffer AVL) prepared with 1µg/µL carrier RNA. The addition of carrier RNA to the reaction mixture reduces degradation of viral nucleic acid by residual RNases and also assists the binding of RNA to the QIAamp Mini column membrane. 560µL of 96-100% ethanol was added and mixed by pulse vortexing. Each sample was passed through a QIAamp Mini column at 6,000 xg for 1 minute. Each sample was then subjected to two wash steps. The first was with 500µL of wash buffer 1 (Buffer AW1) followed by centrifugation at 6,000 x g for 1 minute and the second with wash buffer 2 (Buffer AW2) followed by centrifugation at full speed for 3 minutes. Residual buffer was removed by repeated centrifugation at full speed for 1 minute. RNA was then eluted in nuclease free water in two 40µL volumes. The use of a double elution step is suggested to increase RNA yield to up to 90%. Extracted RNA was either used in downstream applications immediately or stored at -20°C or -45°C.

2.2.3 Reverse transcription

In the majority of studies, a two-step RT-PCR protocol was used. Therefore, single stranded cDNA was generated by reverse transcription (RT) to allow amplification of extracted viral RNA by PCR. RT was carried out using the A3500 Reverse Transcription System (Promega, UK) according to the manufacturers' instructions but with certain modifications.

Briefly, 5µL of eluate obtained from RNA extraction was incubated at 70°C for 10 minutes. 15µL of master-mix containing RT reagents (Table 2.1) was added to each sample. The reaction mixture was then incubated for 10 minutes at room temperature followed by an elongation step which was increased to 50 minutes at

42°C. Samples were then incubated for 5 minutes at 95°C and kept on ice for a further five minutes, in order to inactivate the reverse transcriptase enzyme and prevent it binding to the single stranded cDNA. cDNA was then either used immediately for PCR reactions or stored at -20°C.

TABLE 2.1: Reagents included in reverse transcription reactions

4µL	25mM MgCl
2µL	10X reverse transcription buffer[5]
2µL	10mM deoxynucleotide triphosphates (dNTPs)
0.5µL	Recombinant RNasin ribonuclease inhibitor (25U/µL)
1µL	100µM random hexamers
0.6µL	*Avian myeloblastoma virus* reverse transcriptase (10U/µL)
4.9µL	Nuclease free H_2O

All positive and negative controls for use in subsequent PCR reactions were included in the RT step to ensure consistency of results. The use of random primers in the RT reaction enabled cDNA generated to be used both for the analysis of multiple genome regions and for studies of other RNA viruses. As an alternative to random primers, an RT reaction utilising the outer PCR primers designed for the desired genome fragment can give increased sensitivity. This technique was used in the analysis of certain samples which proved challenging to amplify with standard procedures. However, as this technique precludes the use of generated cDNA for other purposes and is therefore not cost-effective, it was only used sparingly.

In the screening of CSF samples for HRV, samples were combined into pools of 10 prior to RT. 10 sequential samples from the CSF archive were selected and 2µL of RNA from each was combined to create a 20µL pool. This was then thoroughly vortex mixed and 5µL used in the RT reaction. The study of HRV incidence in CSF samples and the repeated screening of certain reported HRV negative respiratory

[5] 10X reverse transcription buffer was supplied with the Reverse Trasncription System kit (Promega, UK) and was composed of 100mM Tris-HCl (pH 9.0 at 25°C), 500mM KCl and 1% Triton® X-100.

samples (described in Chapter 3) were the only studies undertaken for which samples were pooled. Although the value of pooling samples for screening large numbers of patients has been previously demonstrated for human metapneumovirus (HMPV) (Gaunt et al., 2009) and coronavirus (HCoV) (Gaunt et al., 2010), the overall incidence of HRV in respiratory samples is too high to give any value to the procedure in investigations of respiratory epidemiology.

2.2.4 PCR and sequencing of HRV positive samples

Within the studies presented in this thesis, PCR had a dual role. Firstly, samples were screened by PCR based methods (Chapter 3) for the presence of HRV and secondly, specific sequence fragments were amplified for studies of genetic diversity, recombination and evolution within the HRV genome.

2.2.4.1 PCR Primer Design

The design of PCR primers for use in the amplification of various genome regions of all three species of HRV required a delicate balance between primer degeneracy (to counteract high genetic diversity within regions) and primer specificity (to avoid amplification of non-target sequences). All PCR primers were designed in-house, either as part of a previous study (Wisdom et al., 2009b, 2009a) or specifically for the studies described. Upon initial design, each primer set was tested for sensitivity using serial 10-fold dilutions of known positive clinical samples[6].

The first step in the design of PCR primers was to obtain all relevant sequences from Genbank. These aligned sequences were then inspected for areas of high sequence conservation around the target regions. Regions of high sequence conservation were

[6] In 2011, known copy number RNA transcripts were developed of several of these positive clinical samples which had been previously used for both positive controls and sensitivity testing (McLeish et al., 2012). After the advent of these RNA transcripts, sensitivity tests were carried out using known copy number samples at 10-fold dilutions.

then analysed and a fragment for the PCR primer was selected according to the following criteria:

- High sequence conservation – as few degenerate bases as possible
- Length of close to 23-25 nucleotides (not less than 16 and not more than 26)
- Melting temperature (Tm) predicted to be close to 60°C
- G+C content of 40-60% with 3' GC clamp where possible
- No self-annealing or hairpin formation

Calculations of Tm and analysis for self-annealing/hairpin formation were carried out in with the aid of the OligoCalc progam (available at http://www.basic.northwestern.edu/biotools/oligocalc.html). Primer pairs for use in the same PCR reaction were selected to have a Tm of within 5 degrees of each other, where possible. The specificity of each primer set was confirmed by analysis with primer BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast) to ensure that primers did not match human or other non-target sequences in the sample.

TABLE 2.2: Codes for individual nucleotides and ambiguous bases used for design of PCR primers

Base Code	**Nucleotides**
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
K	G or T
M	A or C
S	G or C
W	A or T
Y	T or C
R	G or A
B	G, C or T
V	G, C or A
D	G, A or T
H	A, C or T
N	A, C, G or T

2.2.4.2 PCR protocols

PCR was used to amplify DNA fragments either for detection (in screening) or sequencing. The PCR protocol consists of a number of temperature dependent steps carried out in a cyclical fashion. The reaction mixture (Table 2.3) is subjected to incubation at 94°C which allows denaturation of the target DNA (either cDNA or first round PCR product). The mixture is then incubated at 50°C to allow primer annealing and then at 72°C for nucleotide extension. This is carried out for 30 cycles as standard, followed by a final extension step at 72°C. Specific primers for each PCR reaction are detailed in Appendix 1.

TABLE 2.3: Reagents used in PCR reactions

25mM $MgCl_2$	4µL
3mM dNTPs[a]	2µL
Sense primer (100µM)	1µL
Antisense primer (100µM)	1µL
Taq polymerase[b] (5U/µL)	0.08µL
Either cDNA or first round PCR product[c]	1 – 2µL
Nuclease free H_2O	Up to 20µL

TABLE 2.4: Cycling conditions used in PCR reactions

Number of cycles	**Temperature**	**Time**
30	94°C	18s
	50°C	30s
	72°C	90s
1	72°C	300s

[a] dNTPs were acquired as separate dATP, dGTP, dCTP and DTTP 25mM aliquots (Promega, UK). These were then diluted to make 3mM stocks for use in PCR.
[b] *Taq* polymerase was *GoTaq*® DNA Polymerase (Promega, UK).
[c] For the first round of PCR, 2µL cDNA was incorporated into the reaction mixture. For the second round of PCR, 1µL of first round product was used.

The reagent concentrations and cycling conditions detailed above (Table 2.3 and Table 2.4) represent the standard conditions used. All protocols with new primers were attempted initially under these conditions. However, individual PCR protocols were optimised prior to use with study samples. This involved using 10-fold dilution series of positive control samples and sequentially varying certain parameters to determine the optimal set of conditions for each PCR reaction. Detection of HRV RNA transcript dilutions with 1000 copies present was targeted. The most

commonly modified parameters included the annealing temperature and dNTP concentration. In several cases, the sensitivity of a PCR protocol was greatly improved by using one of the same primers in both rounds (a hemi-nested reaction). All modifications to the standard protocol used for individual PCRs are given below (Table 2.5), along with a reference to the thesis chapter to which these reactions relate.

TABLE 2.5: Modifications to standard PCR protocols used for individual PCR reactions

PCR Reaction	**Name of primer set**	**Modifications to standard reagents**	**Modifications to standard cycling conditions**	**Chapter Reference**
5'UTR Screening	UTR Screen	None	Annealing temperature of 53°C in the second round	Chapter 3
VP4/VP2 Typing	VP4/VP2	None	None	Chapter 3
HRV-A VP1 Amplifcation	A VP1	Specific inner antisense primer used for HRV-A28 positive samples Same first round product used for both second round fragments	None	Chapter 4 Chapter 6
HRV-B VP1 Amplification	B VP1	Hemi-nested reaction – same sense primer in both rounds	Annealing temperature of 48°C in the second round	Chapter 4 Chapter 6
HRV-C VP1 Amplifcation	C VP1 PCR 1	None	None	Chapter 4 Chapter 5 Chapter 6
HRV-C Modified VP1 Amplification	C VP1 PCR 2	All reaction components scaled up to give a final volume of 50µL	Annealing temperature of 45°C in the second round	Chapter 4 Chapter 5 Chapter 6
HRV-C 5'UTR Segment 1	UTR 2	Hemi-nested reaction – same sense primers in both rounds	Annealing temperature of 53°C in the second round	Chapter 5

HRV-C 5'UTR Segment 2	UTR 3	Hemi-nested reaction – same sense primers in both rounds	Annealing temperature of 53°C in the second round	Chapter 5
HRV-C 5'UTR 5' end	UTR 1	Combined amplification of HRV-A/-C used multiple primers in each round (listed in Appendix 1)[a] Hemi-nested reaction – same sense primers in both rounds	Annealing temperature of 53°C in the second round	Chapter 5
HRV-A 3Dpol amplification	A 3Dpol	None	Annealing temperature of 45°C in the second round	Chapter 6
HRV-B 3Dpol amplification	B 3Dpol	None	Annealing temperature of 45°C in the second round	Chapter 6
HRV-C 3Dpol amplification	C 3Dpol	None	Annealing temperature of 45°C in the second round	Chapter 5 Chapter 6
HRV-A 2A amplification	C 2A	Same first round product used for both second round fragments 4µL 3mM dNTPs used in first round[b] Fragment 2 (labelled C2A PCR 2) obtained with a heminested reaction using OAS primer	Annealing temperature of 46°C in the second round	Chapter 5

[a] For reactions using more than two primers, the volume of nuclease free H_2O was adjusted accordingly.

[b] Where volume of dNTPs used is increased, the volume of nuclease free H_2O was adjusted accordingly.

2.2.4.3 SuperScript One-Step reverse transcription polymerase chain reaction (RT-PCR)

For samples which were difficult to amplify using the optimised *GoTaq* protocols listed above, a more sensitive one-step combined RT and PCR protocol was used. This utilised the Superscript III One-Step RT-PCR system with Platinum Taq DNA polymerase (Invitrogen, UK) with parameters and cycling conditions as previously optimised for the amplification of EV (McWilliam Leitch et al., 2009a).

TABLE 2.6: Reagents used in Superscript III RT-PCR reactions

2X Reaction Mix[7]	10µL
Outer Sense PCR primer (100µM)	1µL
Outer Antisense PCR primer (100µM)	1µL
Superscript II reverse transcriptase [8]/Platinum Taq	0.4µL
Extracted RNA	6µL
Nuclease free H_2O	1.6µL

TABLE 2.7: Cycling conditions used in Superscript III RT-PCR reactions

Number of cycles	**Temperature**	**Time**
1	43°C	1 hour
20	53°C	60s
	55°C	60s
1	70°C	15 min
1	94°C	2 min
40	94°C	30s
	50°C	30s
	68°C	90s
1	68°C	5 min

All reactions were prepared on ice and contained reagents as detailed (Table 2.6). RNA directly extracted from clinical specimens was used. As the outer PCR primers for the particular reaction were used for the RT step, the resulting product could only be used for amplification of that specific segment. The Platinum Taq used in the amplification is complexed with an activity blocking antibody that renders it unreactive at room temperature. This prevents non-specific annealing and therefore

[7] 2X Reaction Mix is supplied with the Superscript III One-step RT-PCR kit and contains 4mM of each dNTP and 3.2mM of $MgSO_4$.

[8] In contrast to standard reverse transcription reactions, RT undertaken with the Superscript III system used Moloney murine leukaemia virus reverse transcriptase (M-MLV-RT), as opposed to AMV-RT.

increases sensitivity. The complexed antibody is denatured at the high temperatures used in the thermal cycling (Table 2.7) and polymerase activity resumes. The first round product obtained from the Superscript III reaction was used directly in second round of PCR, as previously described.

2.2.4.4 Gel Electrophoresis

The presence of PCR positive samples and the approximate length of DNA fragment amplified was visualised by agarose gel electrophoresis. 2% agarose gels were prepared with 1X tris-acetate-EDTA (TAE) buffer (Severn biotechnologies, UK) and 1/10000 ethidium bromide (10mg/μL)[9]. 10μL of second round PCR product was thoroughly mixed with 1.5μL of 6X loading dye and loaded directly into a well of prepared gel in an electrophoresis tank containing 1X TAE buffer. As the loading dye is significantly denser than the TAE buffer, it allows the sample to sink into the well. The dye is also negatively charged and so co-migrates with the DNA within the sample, allowing monitoring of the progress of the sample along the gel. An exACTGene Low Range Plus DNA ladder (Fisher Scientific, UK) was loaded alongside the PCR products.

All second round PCR products were electrophoresed at 150 volts for 30 to 45 minutes, depending on the size of the expected product. Resulting bands were visualised using a UV transilluminator.

2.2.4.5 Gel extraction and purification of amplified products

In cases where multiple products were visualized on an agarose gel and the PCR reaction could not be further optimised, bands were excised and purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Life Sciences, UK), according to the manufacturers' instructions. Briefly, DNA bands were manually

[9] Ethidium bromide was used for the majority of initial screening studies undertaken during this thesis. However, as it is an intercalating agent and is a known carcinogen with prolonged exposure, certain University areas discourage its use. For later studies, SybrSAFE was used as an alternative.

excised under UV transillumination with minimal exposure time to prevent DNA damage. Excised bands were incubated at 60°C with 10µL of Capture Buffer 3 per 10mg agarose, until the agarose was completely dissolved. The DNA within the sample was bound to a silica membrane with a Illustra GFX Spin Column by spinning at 16,000 xg for 30 seconds. The sample was then washed with an ethanol containing buffer to remove salts and impurities. Finally, each sample was incubated at room temperature for one minute with 50µL nuclease free water and eluted by spinning at 16,000 xg for 30 seconds. If not required for immediate use, both excised bands and eluted DNA could be stored for up to one week at -20°C.

2.2.4.6 PCR clean-up procedures

All PCR products which were not extracted directly from agarose gels underwent an additional PCR clean-up step to remove unincorporated primers and dNTPs and to improve the fidelity of sequencing reactions. This step was carried out using EXOsap-IT (Fisher Scientific, UK), which is a combination of Exonuclease I to degrade single stranded DNA and Shrimp Alkaline Phosphatase to dephosphorylate excess dNTPS. 2µL of EXOsap-IT was incubated with 5µL second round PCR product at 37°C for 15 minutes. The active enzymes were then denatured at 80°C for 15 minutes. The resulting mixture in theory contained only PCR product, nucleosides and inorganic phosphate and was then used directly for sequencing reactions.

2.2.4.7 Sanger sequencing

All sequencing undertaken during the course of this work employed the Sanger method, whereby the reaction mixture contains all four chain terminating dideoxynucleotides labelled with distinct fluorescent dyes allowing the DNA sequence to be determined sequentially. This was performed using the ABI BigDye Terminator kit (Applied Biosystems, Warrington, UK), with reagents and cycling conditions described below (Table 2.8 and Table 2.9). PCR products were

sequenced in both sense and antisense orientation using second round PCR primers. The amount of BigDye used in each reaction was adjusted according to fragment size, with up to 2µL being used for the largest sequence fragments.

TABLE 2.8: Reagents used in sequencing reactions

PCR product	2µL
BigDye v3.1	1µL - 2µL
Primer (100µM)	1µL
Nuclease free H_2O	to 20µL

TABLE 2.9: Cycling conditions used in sequencing reactions

Number of cycles	**Temperature**	**Time**
25	95°C	30s
	50°C	20s
	60°C	4 min
1	68°C	5 min

Reactions products were sent to the in-house sequencing facility at The Genepool (Ashworth Laboratories, Kings Buildings, Edinburgh) and results were returned in the form of FASTA and corresponding ABI files.

2.3 Computational Methods

2.3.1 Statistical methods

Unless otherwise stated, all statistical tests carried out on epidemiological data were two-tailed Fishers exact tests, appropriate for small datasets with limited categories for each variable.

2.3.2 Sequence alignment

All phylogenetic and evolutionary analyses were entirely dependent on the construction of a sequence alignment in which homologous sequences were correctly identified and aligned. The alignment process is central to all downstream analysis and prevents false conclusions regarding phylogeny, genetic diversity and recombination.

Sequences were imported into SSE v1.0 (Simmonds, 2012) and initially aligned with a combination of the CLUSTAL-W or MUSCLE version 3.8 algorithms (Thompson et al., 1994; Edgar, 2004) implemented within SSE v1.0. Alignments were refined where absolutely necessary by manual editing at the amino acid level (for coding sequences) and nucleotide level (for non-coding sequences). PCR products were sequenced in both sense and antisense directions to allow for dual coverage. Any gaps and mismatches between the sense and antisense sequences were resolved by inspection of accompanying chromatograms in the Chromas software (www.technelysium.com.au). Amino acid sequences were obtained by translation of nucleotide sequences carried out in the SSE v1.0 software package (Simmonds, 2012) and using a standard genetic code.

2.3.3 Database searching

A critically important step in many of the analyses within this thesis was the construction of comprehensive datasets containing all published sequences of a particular gene region or species. The inclusion of all available published sequences with sequences generated from our studies allowed the most comprehensive investigations possible to be carried out. All database searches were carried out of nucleotide sequences available in the three major online databases (Genbank NCBI, European Molecular Biology Laboratory (EMBL) and DNA Databank of Japan (DDBJ)). The existence of a reciprocal agreement between the three major databases to cross-submit all sequences allowed all information required to be accessed via the NCBI Basic Local Alignment Search Tool (BLAST).

The BLAST method of searching sequence databases was first described in 1990 (Altschul et al., 1990) and is based on locating regions of local similarity between a query sequence and all database entries. This procedure was then extended in 1997 to an algorithm more based in biological reality due to its ability to deal with gaps in sequence alignments (Altschul et al., 1997). BLAST is an heuristic, which while fast

and reasonably accurate does sacrifice some of its accuracy for speed and cannot guarantee that all homologous sequences will be returned. The optimization of certain parameters is therefore vital to the accuracy of the endeavour.

All studies presented within this thesis utilised a nucleotide BLAST procedure to obtain the maximum number of matching and potentially homologous sequences for analysis. In brief, an input query sequence in FASTA format is broken down into "words" of a certain length. The entire database is then searched for occurrences of every possible word generated from the query sequence. These have a default length of 11 nucleotides. However, as an exact match of the entire word length is required for progression of the algorithm, the word length can be specified in order to alter the sensitivity and specificity of the protocol. When a match is located in a database sequence, the hit is extended by adding bases from the query sequence at both the 5' and 3' ends and checking for matches in the database sequence. The growing alignment is sequentially scored based on a specified match/mismatch penalty. The default value for match/mismatch score is 1/-2. However, this was decreased to 1/-1 for searches for sequences which were more distantly related. The alignment extension is continually scored until the score drops below a specified threshold (20 for nucleotide sequences). The extension step allows the differentiation of meaningful matches from random ones. Sequences returned were assigned an E-value, which is an estimate of the probability of a hit being a false positive.

In searches which were conducted in order to obtain the total number of VP4/VP2 sequences, specificity was sacrificed for sensitivity in terms of both match/mismatch scoring and word length. Alignment results were then edited manually to contain only sequences that covered >90% of the query region (therefore excluding small, random matches) and those which were annotated as other *Picornavirus* species. All published sequences obtained by database searching are catalogued in our alignments

and presented in phylogenetic trees and other analyses by accession number; which is consistent across all three major databases.

A large number of HRV polyprotein sequences (which are not yet associated with publications) contain large gaps within the coding sequence, presumably due to incomplete coverage during Illumina sequencing. As they are all nearly complete genomes, these sequences were considered as complete genomes and were only excluded if the gaps affected the regions under investigation and comprised of more than 10% of the region.

2.3.4 Likelihood mapping analysis

In the analysis of the occurrence of recombination between three coding regions of HRV-A, -B and –C, likelihood mapping analysis was used to visualise the phylogenetic signal content of the alignment *a priori* (Strimmer and von Haeseler, 1997). In addition, this method confirms that sequences have undergone the correct mode of evolution to result in a resolved bifurcating phylogenetic tree and therefore that conclusions based upon such an inferred tree would be valid.

Likelihood mapping analysis is based on quartet puzzling and was implemented in the TREE-PUZZLE v5.2 software (Schmidt et al., 2002). Briefly, maximum likelihoods were calculated for each of the 3 possible unrooted trees derived from 4 randomly chosen sequences from the alignment (a quartet) (Strimmer, 1994). Using the Hasegawa-Kishino-Yano (HKY) model of substitution, a total of 10,000 quartets were analysed and each was assigned into a particular category based on the maximum likelihood. The results are plotted onto a triangular surface, whereby the proportion of quartets in each of the three vertices represent fully resolved bifurcating phylogenetic trees and those located in the sides of the triangle represent equal observed likelihoods between two of the calculated trees. The proportion of quartets in the centre of the triangle represent those with an unresolved star-like phylogeny, whereby the likelihoods of all three possible trees are equal. Any

analysis with more than 20-30% of analysed quartets showing unresolved phylogeny are not considered reliable for phylogenetic reconstruction. Possible explanations include "noisy" data, lack of sequence homology, alignment errors and recombination within the region subjected to the analysis.

Likelihood mapping analysis has been previously used to confirm the suitability of a subset of HRV VP4/VP2 sequences (Savolainen et al., 2002) and EV sequences (Andersson et al., 2002; Lindberg et al., 2003) for phylogenetic analysis. In addition, the method has also been used to demonstrate the star-like bursting radiation resulting from the rapid transmission of human immunodeficiency virus 1C (HIV-1C) after introduction to South America (Véras et al., 2011) and to assess suitability of different regions of HAV for genotyping (Joshi et al., 2008).

2.3.5 Construction of phylogenetic trees

Phylogenetic analysis uses sequence information to provide an estimate of the evolutionary relationships between sequences, provided that adequate homology exists. Studies within this thesis were largely reliant on the use of the Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (Tamura et al., 2007) and latterly, version 5.0 (Tamura et al., 2011). Unless otherwise stated, all phylogenetic trees used for HRV typing and analyses of recombination were constructed by neighbour-joining method (Saitou and Nei, 1987) from 100 bootstrap re-sampled sequence alignments of maximum composite likelihood (MCL) (Guindon and Gascuel, 2003) distances with pairwise deletion for missing data [10]. Phylogenetic trees were improved by the inclusion of an outlier, which rooted the tree and therefore provided an estimate of which nodes under study were the most ancient. For full HRV species analyses that were conducted, sequence NC_001430: EV70 was used to root the tree. However, for analyses of less genetically diverse

[10] Phylogenetic trees included in Chapter 6 were latterly constructed with 1000 bootstrap repetitions. These were compared directly with trees constructed with 100 bootstraps and no detectable difference was found.

sequences (within HRV species), a sequence of different species or type was used (specified throughout).

The neighbour-joining (NJ) method was first described by Saitou and Nei in 1987 (Saitou and Nei, 1987). It is a distance based method that produces an additive tree and is one of the most widely used methods for phylogenetic tree reconstruction (Gascuel and Steel, 2006). While the authors have stated that the method is not guaranteed to always produce the minimum evolution tree, it has good reliability in producing the correct tree topology (Saitou and Nei, 1987). The fact that it is significantly more computationally efficient than more traditional methods based on maximum parsimony allows the relatively fast calculation of trees reconstructed from a large number of operational taxonomic units (OTUs). As the name suggests, NJ is based on the principle of sequentially joining the most closely related OTUs via a single interior node in an unrooted bifurcating tree. Initially, all OTUs/sequences are represented in a star-like tree with no internal clustering. Evolutionary distances are calculated in a pairwise fashion between all sequences and the OTU pair with the shortest inferred branch lengths are joined by an internal tree node. The algorithm then considers this pair as a single OTU and repeats the process of analysing evolutionary distances between all sequences, culminating in the joining of the pair to their nearest neighbour. This process is repeated sequentially until all sequences are included in the final tree.

One of the limitations of the NJ method is that it assumes that no backwards or parallel substitutions have taken place. Therefore, NJ trees are subject to long branch attraction (Felsenstein, 1978), whereby it can predict that two OTUs are more closely related than they actually are, as the algorithm cannot correct for multiple substitutions at specific sites. Other phylogenetic reconstruction methods such as those relying on maximum likelihood optimality criterion are generally more biologically accurate due to their greater ability to correct for multiple hits (Holder

and Lewis, 2003). However, as maximum likelihood methods of tree-building seek to explore a significant proportion of the tree space and score each tree, they are very computationally intensive methods and not practical for use on large datasets.

The dual problems of inaccurate phylogeny estimation and long branch attraction exhibited in pure NJ algorithms can be prevented by the usage of an MCL method to calculate evolutionary distances within a dataset (Tamura et al., 2004). The calculation of MCL distances is ideal for rapidly evolving sequences, as it automatically corrects for multiple mutational events at the same site and additionally incorporates a model of nucleotide substitution which is optimized automatically to fit the data (Tamura et al., 2011). This includes such aspects as transition/transversion ratios, whereby it is recognised that purine-purine and pyrimidine-pyrimidine substitutions occur more readily than purine-pyrimidine in natural mutations.

For datasets that were analysed with Bayesian Evolutionary Analysis of Sampling Trees (BEAST) software (Drummond and Rambaut, 2007), phylogenetic trees were additionally constructed by Bayesian methods implemented in the program. These were found, in each case, to present identical tree topology to trees produced by the neighbour-joining maximum composite likelihood method.

2.3.6 Bootstrapping

In order to determine the robustness of observed phylogenetic groupings within the tree, all trees were constructed with either 100 or 1000 bootstrap resamplings. The bootstrap is a statistical technique that was first applied to assessing confidence levels in phylogenetic groupings by Professor J. Felsenstein of the University of Washington in 1985 (Felsenstein, 1985). The generation of bootstrap values for internal nodes of a phylogenetic tree gives an estimate of the reproducibility of a monophyletic clade. Sequence data used to construct the original tree is randomly resampled and trees constructed from each of the pseudo-replicates produced. The

proportion of observed trees in which each monophyletic clade is observed is then computed and reported as the bootstrap value (Holder and Lewis, 2003). A high bootstrap value (close to 100) indicates that analysis of randomly selected subsets of the alignment data consistently predict the same clade (Berry and Gascuel, 1996). One of the major limitations of bootstrap analysis is that it fails to provide an estimate of the probability of an observed tree and merely indicates a degree of robustness for the tree phylogeny. For instance, due to the inherent bias in methods of calculation of evolutionary distances, long branches subject to long branch attraction can have good bootstrap support, even if they do not represent a "true" phylogenetic relationship (Swofford et al., 2001).

For all phylogenetic studies included in this thesis, a bootstrap value of 70 was used as the threshold for determining adequate support. MEGA v4.0 and v5.0 software packages present phylogenetic trees as both a majority rule bootstrap consensus tree and as the originally determined tree with bootstrap values superimposed. The bootstrap consensus tree was used for phylogenetic analyses.

2.3.7 Recombination detection methods

The general strategy for the detection of recombination in sequence alignments was multi-faceted. Central to all analysis undertaken was the topology of bootstrap supported phylogenetic trees. Phylogenetic trees were constructed containing sequence fragments of varying sizes (depending on the specific analysis undertaken). Trees from sequential genome regions were manually inspected to identify phylogeny violations between clades with bootstrap support above the threshold of 70. All recombination events detected from the inspection of phylogenetic trees were further characterised by analysis with GroupScan and where possible, putative breakpoints were determined.

In addition to the initial screening of sequence datasets by phylogenetic reconstruction, several computational analyses were undertaken to effectively screen

each dataset for the occurrence of recombination. This was done to supplement and support results obtained from phylogenetic analysis. Although a great number of methods for the detection of recombination in sequence data currently exist, methods often give differing or even, contradictory results (Posada and Crandall, 2001; Posada, 2002) and therefore a variety of different approaches were combined within this work. Screening analyses were undertaken with programs implemented in the Recombination Detection Program (RDP) v4.0 (Martin et al., 2010) and SSE v1.0 (Simmonds, 2012) software packages and those available through the HyPhy datamonkey webserver (www.datamonkey.org). All algorithms used for the detection and description of putative recombination events are discussed below.

2.3.7.1 Recombination Detection Program v4.0

The initial screening of each sequence dataset for the occurrence of recombination with bioinformatics based methods was undertaken using algorithms implemented in RDP v4.0 (Martin et al., 2010). RDP v4.0 incorporates numerous different methods for the analysis of recombination in sequence data. Of these, we subjected our datasets to analysis with RDP (Martin and Rybicki, 2000), GeneConv (Padidam et al., 1999), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001), SiScan (Gibbs et al., 2000) and BootScan (Martin et al., 2005). A putative recombination event was accepted for further analysis if it was detected by more than two of these methods and returned a p-value of less than 0.05. Although these algorithms employ varying methods of detecting and characterising potential recombination events, only BootScan within the RDP package takes account of phylogenetic incongruence observed between different sequence fragments. As GroupScan offers a preferable method to BootScan for the identification of putative recombination breakpoints (Section 2.3.7.4), the algorithms available within the RDP v4.0 software package were used only for screening purposes.

2.3.7.2 Genetic Algorithm for Recombination Detection (GARD) and Single Breakpoint Recombination (SBP)

All coding region datasets analysed for the occurrence of recombination were analysed with the Genetic Algorithm for Recombination Detection (GARD) and Single Breakpoint Recombination (SBP) programs (Kosakovsky Pond et al., 2006b, 2006a), available on the HyPhy Datamonkey webserver (www.datamonkey.org).

The two programs operate in a highly similar manner. However, SBP seeks to confirm the presence of recombination within a sequence alignment and GARD seeks to locate all potential recombination breakpoints. The method is based on a search for sequence specific phylogenies between different regions of an alignment, concentrating on a single inferred breakpoint for SBP and with a more heuristic approach in GARD. Briefly, a NJ tree is constructed for the full sequence alignment and an AIC_c (small sample Akaike Information Criterion) score is calculated. The score takes into account the nucleotide model specified and rate parameters and branch lengths are estimated by maximum likelihood. The sequence alignment is then split into sequential blocks with a breakpoint (i) in the middle. An NJ tree and corresponding AIC_c score are calculated for each block. In SBP, this process is repeated for every possible location of the breakpoint. If the AIC_c score is higher for any individual segment than for the alignment as a whole, recombination is deemed to have acted on the alignment. In GARD analysis, a more heuristic approach to searching the space for putative breakpoints is employed, as scanning every possible combination of breakpoint locations is not practical. In order to infer a putative recombination breakpoint, the algorithm does not always require a topological change to be observed. Instances of changes in branch length between segments are additionally reported (Kosakovsky Pond et al., 2006b, 2006a).

Datasets were analysed both as full HRV species and as individual HRV types. In addition, sets of sequences involved in recombination events observed by other

methods (ie/ RDP or phylogenetic incongruity) were specifically subjected to analysis with GARD and SBP, in order to verify events observed. Due to the restriction placed upon individual server usage, a full analysis of all three HRV species concurrently was not able to be carried out in this manner. The model selection tool offered within the datamonkey webserver was employed and the appropriate model of nucleotide substitution was used for each GARD/SBP analysis.

2.3.7.3 TreeOrderScan

The program TreeOrderScan in SSE v1.0 is a sequence analysis tool used to directly visualize the frequency of recombination events in a particular sequence alignment and provide a preliminary analysis of potential breakpoints. TreeOrderScan analyses a sequence alignment over sequential fragments of defined length and increment. Optimal trees are produced for each sequence fragment and branches collapsed to a pre-specified bootstrap support threshold. The program then directly compares the branching order of trees from each fragment and catalogues any phylogeny violations. This is then presented as a half-diagonal matrix, giving an overview of the degree of phylogenetic incongruence observed in each region and additionally as a specific list of putative violations. Sequence datasets can be sub-divided into groups, which then allows the computation of frequencies of phylogeny violations both within and between groups (for instance, different HRV species).

TreeOrderScan was used as an initial tool to screen alignments for recombination, using a sequence fragment of 300 nucleotides and an increment of 30. This was implemented in analysis of recombination within the 5'UTR of HRV-C (Chapter 5) and analyses of recombination within the coding regions of all three HRV species (Chapter 6). All analyses were undertaken using EV70 : NC_001430 as an outgroup. This technique has also been used for the analysis of recombination in apthoviruses (Heath et al., 2006), HRV (Linsuwanon et al., 2011) and EV (Simmonds and Welch, 2006).

2.3.7.4 GroupScan

In instances where recombination was detected by inspection of phylogenetic trees and the usage of the computational methods described above, GroupScan was used in order to verify putative events and determine recombination breakpoints where possible. GroupScan is also implemented within the SSE v1.0 software package.

GroupScan was first described in 2005 (Simmonds and Midgley, 2005) and is used to score the extent of grouping of a query sequence with two or more control groups. The grouping score is a measure of how deeply embedded within a particular group a query sequence is. For instance, a score of 0 indicates that there is no grouping and a score of 1 indicates that the query sequence groups deep within a clade of control sequences. Sequences are analysed in a sequential fashion, with both fragment size and increment being specified by the user. The output shows a graph with the midpoint of the analysed fragment represented on the x axis and the grouping score on the y. This can be used to determine approximate recombination breakpoints, when a query sequence is scanned against its nearest neighbours in both regions. The putative breakpoint is determined by point of intersection of the grouping score trace of both parental sequences.

The GroupScan method was used for all determinations of recombination breakpoints. In preference to the method commonly employed in BootScan and Simplot analyses, GroupScan can consider entire groups of sequences as control groups and therefore doesn't lose valuable phylogenetic information by condensing these sequences to a single consensus. This also allows the calculation of recombination breakpoints in circumstances where parental sequences are not present within the dataset analysed. The RDP algorithm also returns estimations of recombination breakpoints. However, the RDP method relies purely on the analysis of evolutionary distance data, whereas GroupScan bases the score on observations conducted of bootstrap supported phylogenetic trees.

Within this thesis, GroupScan was used to pinpoint recombination breakpoints both in coding (Chapter 6) and non-coding (Chapter 5) genome regions. The program was also used as an adjunct to other analyses in the further characterisation of putative recombination events highlighted by computational analyses, but not observed on bootstrapped phylogenetic trees. All analyses carried out used a fragment size of 300 nucleotides and an increment of 30, unless otherwise stated. Coding region recombination breakpoints (Chapter 6) were additionally analysed with a 600 base fragment and 30 base increment. These additional program runs generated highly similar results.

2.3.8 Bayesian Evolution Analysis of Sampling Trees (BEAST)

In order to model the evolutionary history of selected HRV sequence fragments and determine mean substitution rates, the BEAST software package was used (Drummond and Rambaut, 2007) to out Bayesian Markov Chain Monte Carlo (MCMC) analysis of input sequence datasets. Sequence data is prepared for analysis using the adjunct software Bayesian Evolutionary Analysis Utility (BEAUti). This allows input of tip dates for sequences (where known) and specification of priors, including a nucleotide substitution model and molecular clock. The SRD06 model was used for all studies within this thesis, as has been recommended for the analysis of protein coding sequence data (Shapiro et al., 2006). This model specifies an HKY model of nucleotide substitution, assumes equal base frequencies and specifies gamma distributed rate heterogeneity. Additionally, the SRD06 model links the first two bases of a codon but allows the third base to have a different substitution rate and transition/transversion ratio. A constant population size prior and a relaxed molecular clock were used for all analyses. A relaxed molecular clock does not assume a constant mutation rate between lineages and so can assign different mutation rates to different tree branches. All other priors are optimized during the execution of the program.

The XML file generated in BEAUti was input into BEAST and the MCMC algorithm was run for 20 million steps. Initially, the program produces a phylogenetic reconstruction from the input sequence data and calculates the posterior probability density of this tree. The algorithm then randomly varies a few of the input parameters to create a small, sequential step in the parameter space. The posterior probability density of this step is calculated and a decision is made regarding the position of the chain in the parameter space. If the posterior probability of the new step is higher, the chain moves on. If a large drop is observed, then the chain maintains its position and an identical step is created. However, if the posterior probability density decreases only slightly, the move is accepted in a proportion of cases. This process is then repeated millions of times. As the steps taken are small and incremental, chains tend to stay in regions of high posterior probability and therefore converge upon a specific reconstruction with a high posterior probability.

BEAST output was analysed in the Tracer program, as part of the BEAST package. On completion, inspection of the output files can help to give an indication of whether the number of steps in the chain was sufficient. Each output parameter is reported with an estimated sample size value (ESS). Values of less than 200 indicate a requirement to optimize input priors or increase the number of MCMC steps. In order to reconstruct phylogenetic trees produced by Bayesian methods, tree must be annotated in TreeAnnotator program (also part of the BEAST package) and visualized in the FigTree software.

In order to verify that the specified priors were not inadvertently influencing the results obtained, each sequence dataset was subjected to an "empty" run of the BEAST algorithm with sequence data excluded. In addition, each sequence dataset was analysed with the dates of isolation of individual sequences scrambled.

Chapter 3
Epidemiology and Clinical Correlations of HRV

3.1 Introduction

3.1.1 Characteristics of circulating HRV in populations of patients with respiratory illnesses

HRV are very common respiratory infections and the classic triad of rhinorrhoea, cough and headache has been experienced on numerous occasions by most of the population. Seropositivity for HRV exceeds 90% by two years of age (Blomqvist et al., 2002a). However, even this staggeringly high seroprevalence is likely to be an underestimate, as the methods employed in this study would have overlooked not only 12 now recognised HRV-A/-B types[11] (www.picornaviridae.com) but the entire species C. The gradual shift to the use of molecular based methods for HRV screening has enabled much more accurate estimates of HRV incidence and worldwide circulation to be produced.

HRVs are routinely detected in respiratory samples in every month of the year and all three species have been found to have a global distribution (Briese et al., 2008). The documented detection frequency of HRV in respiratory samples ranges from 7.9% - 64.3% (Alper et al., 2008; du Prel et al., 2009; Brittain-Long et al., 2010; Chan et al., 2012; Chidlow et al., 2012; Renwick et al., 2007; Lamson et al., 2006; Arden et al., 2006; Kistler et al., 2007a)[12]. Two seasonal peaks in HRV detection are generally observed; one in the early autumn and one in the spring (Jin et al., 2009;

[11] Throughout this chapter, the term "HRV type" is used to refer to a phylogenetic group of HRV sequences with pairwise nucleotide p-distance in VP4/VP2 of below the proposed threshold for each species (Simmonds et al., 2010; McIntyre et al., 2013a). The process of defining these thresholds and assigning HRV sequences into genotypically assigned types in described in Chapter 4.

[12] Only detection rates from published studies which used RT-PCR based screening methods are given.

El-Sahly et al., 2000). In contrast to other common respiratory viruses and other EV species, HRV types do not appear to circulate in an epidemic fashion (with a small number of circulating types being periodically replaced). Instead, a large number of genetically distinct HRV types circulate concurrently (Savolainen et al., 2002; Wisdom et al., 2009b), with no discernible predominance of a particular strain or species (Fry et al., 2011). The existence of over 150 distinct types and the lack of cross-immunity make this constant circulation possible. Despite the abundance of HRV types circulating during one peak season, there appears to be relatively limited persistence of types from one season to the next. One study which considered HRV circulating between September 2006 and Feburary 2007 found that only 5 of the 43 types observed were present in both study months and one HRV-C positive patient was found to be re-infected with HRV-A only 5 days later (Wisdom et al., 2009b). The characteristic, almost random, fashion in which HRV appears to circulate in respiratory samples makes discerning patterns and possible networks of transmission nearly impossible.

HRV-A strains are generally the most frequently detected, while HRV-B is comparatively rare (Chidlow et al., 2012; Wisdom et al., 2009b). A few sporadic studies, mostly originating from the Asian-Pacific region have reported a higher prevalence for HRV-C (Chan et al., 2012; Jin et al., 2009). No consistent bias towards infection of male or female patients has been evident in published studies. Although one study found an a slight excess of HRV-C infections in female patients (Wisdom et al., 2009b), another found a slight excess overall of HRV in male patients (Chan et al., 2012). Overall, precise epidemiological characteristics of HRV infections are difficult to discern, due to the very high incidence in the general population and the large number of distinct strains circulating concurrently.

3.1.2 Involvement of HRV in infections out with the upper respiratory tract

The classical perception of HRV as merely an agent of the "common cold" caused this group of viruses to be largely overlooked for many years. In addition to their known association with mild URTIs, the realisation that HRV can effectively replicate in the lower airways has led to a now extensive association with lower respiratory tract infection (LRTI), especially in immunocompromised patients (Gutman et al., 2007; Parody et al., 2007; Papadopoulos et al., 2000; Kusel et al., 2006; Papadopoulos et al., 1999; Han et al., 2009; Linsuwanon et al., 2009; Garbino et al., 2009). The reported detection frequency of HRV in lower respiratory tract specimens ranges from 12.6% to 37.8% (Fry et al., 2011; de Almeida et al., 2010; Costa et al., 2011; Arden et al., 2006). HRV are also clinically important causes of exacerbation of chronic lung diseases, such as asthma (Khetsuriani et al., 2008; Bizzintino et al., 2011; Miller et al., 2009a; Friedlander and Busse, 2005) and cystic fibrosis (de Almeida et al., 2010; Smyth et al., 1995).

In addition to severe illness within the respiratory tract, HRV has been implicated in the development of acute otitis media (AOM) (Nokso-Koivisto et al., 2006; Pitkäranta et al., 1998; Savolainen-Kopra et al., 2009a). Indeed, one study isolated HRV from middle ear fluid in 41% of episodes of AOM (Blomqvist et al., 2002a). However, this rate was obtained from screening samples by cell culture based methods and is therefore likely to represent a significant underestimate of the true incidence. HRV have also been linked to the development of acute and chronic sinusitis (Pitkäranta et al., 2001, 1997), which are significant causes of morbidity in the general adult population.

HRV have also been isolated, albeit infrequently, from blood samples of infected patients. This is suggestive of a potential role of HRV viraemia in development of severe disease (Xatzipsalti et al., 2005). Several isolated case reports have discussed

the occurrence of HRV in critically and systemically unwell patients, with HRV RNA being obtained from such distant sites as plasma, pericardial fluid, urine and stool (Broberg et al., 2011; Tapparel et al., 2009c). However, these reports of isolation of HRV from sites other than the upper and lower respiratory tract do remain rare. The current perception of HRV as predominantly causative agents of mild and unimportant illness has probably stifled the desire of the research community to investigate other potential clinical associations.

3.1.3 The role of enteroviruses in aseptic meningitis

The four known species of EV (EVA-D) are often isolated from faecal samples and have been associated with a wide variety of acute and chronic clinical illness (Pallansch and Roos, 2007). However, despite a broad spectrum of clinical associations, all four species have been linked with CNS infection, especially aseptic (viral) meningitis.

Although viral meningitis is generally self-limiting, it can potentially lead to devastating sequelae for affected patients, including seizure disorders and rapid deterioration of mental state in cases of progression to meningoencephalitis (Valcour et al., 2008). A viral aetiology can only be defined in around 66% of cases of aseptic meningitis and of these EV are the most common, with an overall detection frequency of around 26% (Kupila et al., 2006). Other clinically important causes of viral meningitis include herpes simplex viruses (HSV) types 1 and 2 and varicella zoster virus (VZV).

Despite the close genetic relationship of HRV with EV and known phenotypic overlap in the causation of respiratory infections, to our knowledge, a potential relationship of HRV with previously undiagnosed cases of aseptic meningitis had not been examined. Interestingly, HRVs have been isolated during the molecular surveillance of sewage samples in Finland, Latvia and Slovakia (Blomqvist et al.,

2009) and in investigation of acute flaccid paralysis in South Asian children (Victoria et al., 2009). The gastrointestinal tract is the major site of EV replication and many cases of meningitis are diagnosed on the basis of EV in stool samples. HRV have been traditionally thought to be unable to infect the gastrointestinal tract secondary to acid lability. The isolation of HRV strains from environmental sewage samples was initially attributed to swallowed respiratory secretions or discarded tissues (Blomqvist et al., 2009). One previous case report detailed the isolation of an HRV-C strain from the stool of a systemically unwell patient with pneumonia and pericarditis (Tapparel et al., 2009c). In this case, HRV was also isolated from blood plasma and pericardial fluid. However, the significance of the HRV positive stool sample was uncertain as it had a low observed viral load. The detection frequency of HRV in stool samples had, at the time of the completion of our study, not been discussed.

The studies presented within this chapter further investigate several intriguing aspects of the epidemiology of HRV. Two studies of the epidemiology of HRV respiratory infections in Edinburgh aimed to further expand upon previous work carried out in this geographical region (Wisdom et al., 2009b) by analysing clinical features of HRV positive patients and directly comparing distribution of HRV species between patients with severe respiratory disease necessitating care in the ITU and those cared for on a general ward-based setting. Preliminary investigations of the occurrence and detection frequency of HRV infections in both CSF and stool specimens were conducted, involving 1500 and 288 referred samples respectively. The results of these studies have highlighted a potential overlap in the disease manifestations of HRV/EV and highlight the requirement for further study of these areas.

3.2 Materials and Methods

3.2.1 Sample selection

3.2.1.1 Sample selection for screening of respiratory samples

In order to investigate the clinical correlations of HRV species and types circulating in respiratory samples in Edinburgh, 556 samples were selected from the SVC archive for screening. These were selected randomly from all samples referred for testing between September 2008 and May 2009.

The SVC archive has been maintained since October 2005. The total number of samples referred to the SVC has increased slightly every year and regular peaks in the number of referrals are additionally observed during the so-called "cold and flu months". These peaks coincide with the Health Protection Agency (HPA) influenza monitoring programme, which runs from October to May (www.hpa.org.uk).

A sharp increase in the number of samples referred to the SVC was noted after the announcement of the H1N1 global swine flu pandemic in June 2009 (Figure 3.1). In particular, the proportion of samples referred from community based patients increased dramatically, from between 1-3% in 2006 – 2008 to 9% in 2009 (marked in black on Figure 3.1). The peak of referrals from community health care centres was 21% of all referred samples in June 2009, coinciding with the official announcement by the World Health Organisation (WHO) of the H1N1 swine flu pandemic. This increase could be attributed to greater public awareness and anxiety regarding cold and flu like illness (CFLI), a lower referral threshold and the practice of routine sampling of asymptomatic contacts in suspected cases.

As a consequence of the artificially increased numbers of referred samples, it was deemed both inappropriate and impractical to use archived samples from this period for this study. Therefore, samples were randomly selected from the 9 month period immediately prior to June 2009. The universally high observed detection frequency

of HRV in respiratory samples diminishes the value of pooling samples for screening and therefore they were selected randomly to provide a cross-section of the time period, while still maintaining manageable numbers for a small study.

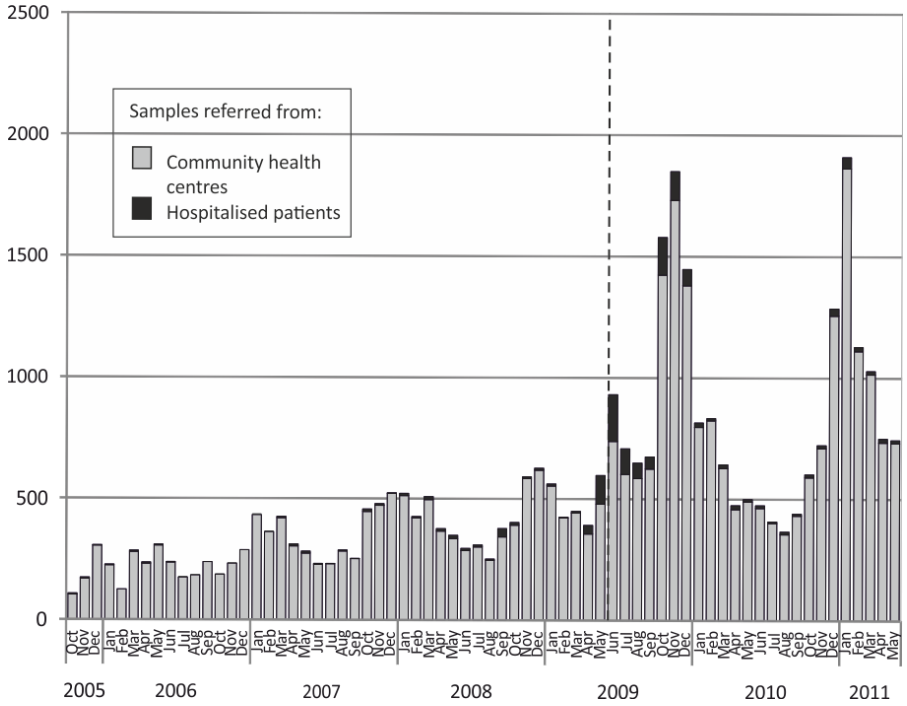


FIGURE 3.1: Total number of respiratory samples referred to the SVC for virological testing per month from October 2005 to May 2011. *Sections coloured in black represent the proportion of samples that were referred from community settings, whereas sections coloured in grey represent those referred from hospitalised patients. The date of declaration of the swine flu pandemic by the WHO is marked by a dotted line.*

3.2.1.2 Sample selection for comparison of patients with acute and critical illness

In order to investigate clinical associations between HRV species and severity of illness, a preliminary study was conducted comparing characteristics of HRV positive patients in both the Intensive Treatment Unit (ITU) and the Accident & Emergency department (A&E) or general wards. A small subset of HRV positive samples from July 2007 until June 2008 with comprehensive patient information available were selected for inclusion. These had previously undergone routine

diagnostic screening for respiratory viruses by staff at the SVC using standard protocols (Bennett et al., 2011; Dierssen et al., 2008; Scheltinga et al., 2005). Samples which were from the same patient were excluded. All other data, including age and clinical details were obscured during the selection process.

3.2.1.3 Sample selection for screening of CSF samples

In parallel with the respiratory archive, an archive is maintained of all referred CSF samples and corresponding extracted RNA. In order to investigate HRV as a potential causative agent of aseptic meningitis, one full year (2009) of referred CSF samples were selected and extracted RNA was combined into pools of 10 for screening.

3.2.1.4 Sample selection for screening of stool samples

The study presented within this thesis involved screening a total of 288 referred stool samples for HRV. This work formed a part of a larger published study that included screening undertaken at the SVC by Dr H. Harvala of a further 333 diagnostic stool samples by real-time PCR (Harvala et al., 2012b). The 288 samples included in this study comprised of stool samples isolated in May (n=90), September (n=108) and December (n=90) of 2010. These represented referred specimens for routine diagnostic analysis and included a number of patients referred in the investigation of suspected norovirus outbreaks.

3.2.1.5 Sample selection for screening of respiratory samples from September and December 2010

In order to determine whether detection of HRV in stool specimens coincided with an increase in the same strains isolated from respiratory specimens, screening of a subset of respiratory samples was undertaken. These were taken from the two calendar months with the highest detection frequency of HRV in stool samples. Due to the concentration of HRV infections in stool in children under the age of five and

older adults, the sample selection was restricted to these age bands. This led to the inclusion of a total of 731 respiratory samples (September: n= 221, December: n=510).

3.2.2 Amplification of 5'UTR and VP4/VP2 region of HRV positive samples

All samples were initially screened for HRV using a nested RT-PCR strategy for the 5'UTR of HRV and positive samples were then amplified in the VP4/VP2 region (described in Chapter 2). VP4/VP2 sequences of samples grouped into phylogenetic clusters (types), in accordance with grouping with an HRV prototype strain.

3.3 Results

3.3.1 Epidemiological characteristics of HRV circulating in Edinburgh between September 2009 and May 2010

556 samples were screened in the 5'UTR for HRV and VP4/VP2 typing was attempted for the 141 detected positive samples. Of these, 8 samples (6%) were found to be EV strains, 6 samples could not be typed and 127 (23 %) were confirmed to be positive for HRV. HRV-A was the most frequently observed (54%), followed by HRV-C (36%) and HRV-B was present in only 10% of samples The observed prevalence of HRV infection was slightly higher in males (26%) than in females (20%) and a broadly similar relative frequency of HRV species was observed in both groups (Figure 3.2 A). HRV was significantly more frequently detected in hospitalised patients than in patients in community based or outpatient settings (p=0.0001). There was additionally no evidence of predominant circulation of any particular HRV species or type within any time period.

Samples included were divided into those obtained from the upper respiratory tract (including nasopharyngeal aspirates) and those isolated from the lower respiratory

tract (including bronchoalveolar lavage). The majority were taken from the upper respiratory tract (n=476; 86%). There was no difference in detection frequency of HRV recorded in upper and lower respiratory samples (23% and 24% respectively). However, there was a statistically significant association between HRV-A detection and isolation of sample from the lower respiratory tract (p=0.034) (Figure 3.2 B). HRV infection was more frequently detected in patients under the age of 5 (31.2%) than in those aged 6 and over (12.1%) (p=0.001). The relative frequency of different HRV species was largely similar across different age bands (Figure 3.2 C), with the only statistically significant difference being a higher rate of detection of HRV-C in the 7-12 month age group (p=<0.0001).

The samples included in this study had previously undergone diagnostic screening at the SVC for adenovirus (AdV), parainfluenza virus (PIV) types 1-4, influenza A (FluA) and influenza B (FluB). In addition, these samples had been previously screened for three types of human coronavirus (HCoV); HKU1, OC43 and 229E (Gaunt et al., 2010). The most common co-infection observed was by AdV representing 10.2% of HRV infections (n=13). There were four triple infections and one quadruple infection, all involving both HRV and AdV. All triple infections originated from paediatric patients for in the general ward. However, the quadruple infection was observed in a 7-12 month old ITU patient. Detailed clinical data for this particular sample is missing, although the sample type is recorded as a bronchoalveloar lavage specimen. Bronchoalveolar lavage is a traumatic and invasive sampling technique which is only carried out when indicated by high clinical suspicion of lower respiratory tract pathology. This supports the assumption that there is likely to have been a relatively severe lower respiratory tract involvement in this child's illness.

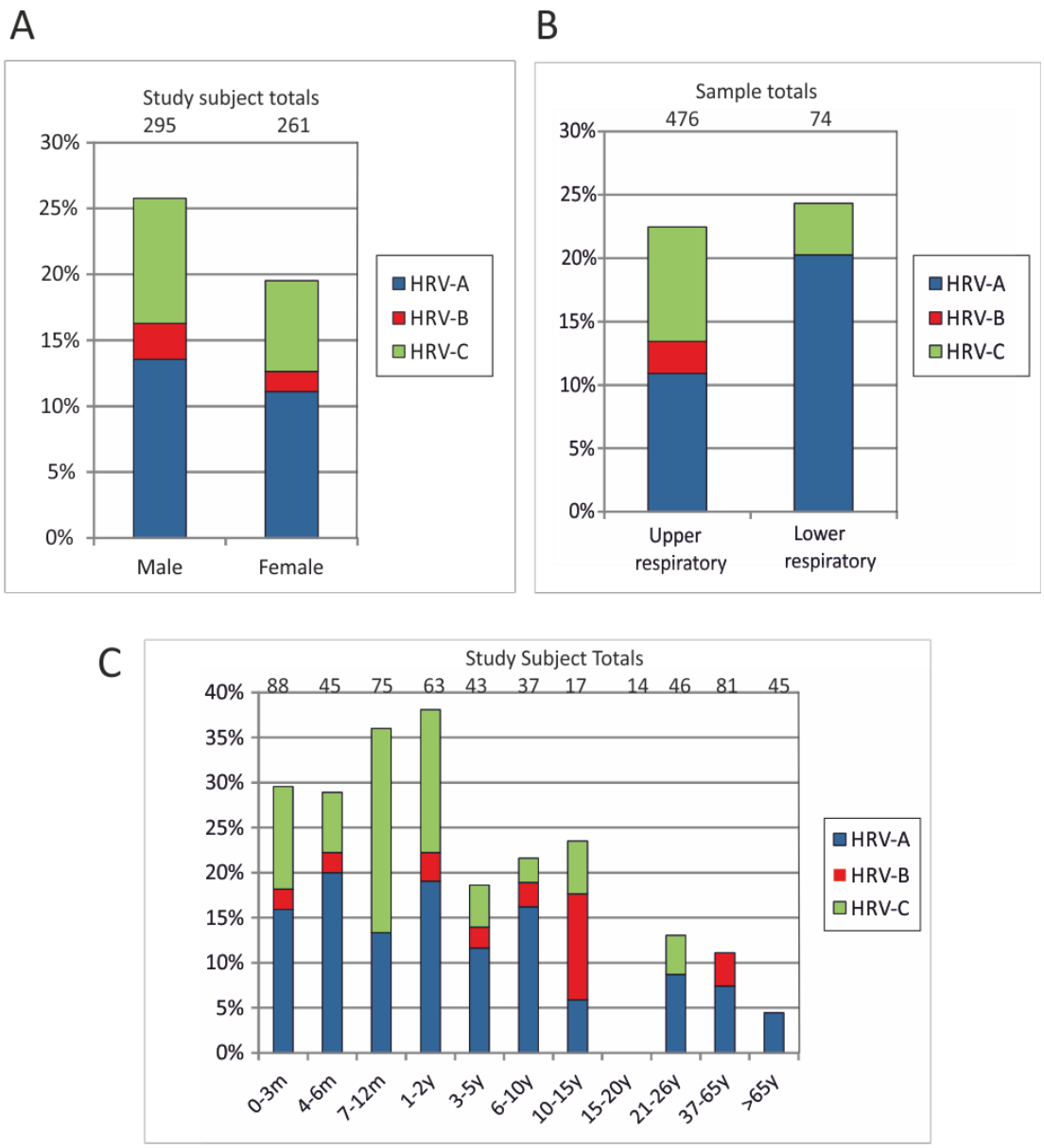


FIGURE 3.2: Epidemiological analysis of HRV infections in patients from September 2008 – May 2009. *Comparison of the detection frequency of HRV-A, -B and –C infections in male and female subjects (A) and upper and lower respiratory tract samples (B). Subject/sample totals are given above each graph. Six samples had no sample type information available. (C) Detection frequency of HRV-A, -B and –C in each age band studied (m = months, y = years).*

One paediatric patient, aged 7-12 months, appeared to harbour a persistent infection with both the same HRV-C (C8) and HCoV (OC43) types over a 2 month period. This patient was suffering *Pneumocystis carinnii* pneumonia, secondary to severe combined immunodeficiency and requiring mechanical ventilation. Besides HRV infection, no other co-infections were noted. These observations suggest that persistence of HRV may occur in immunocompromised patients who are critically unwell.

3.3.2 Comparison of characteristics of HRV positive paediatric patients between acute and high dependency care settings

In order to investigate clinical characteristics of HRV positive paediatric patients in acute and high dependency care settings, a subset of 64 samples were typed in the VP4/VP2 region. These were all from patients being treated in the Royal Hospital for Sick Children, Edinburgh (RHSC) and spanned a single calendar year from July 2007 to June 2008. 35 samples (55%) originated from patients being cared for in a high dependency ward setting, including both the ITU and high dependency unit (HDU). 29 samples (45%) were taken from patients during acute and short-term admissions. Acute admissions were defined as those patients discharged from A&E or those under short-term observation within the Medical Assessment Unit (MAU). This group had an average length of stay in hospital of 2.42 days (range 0 to 13 days), whereas high dependency patients had an average length of stay in hospital of 35 days (range 3 to 170 days).

Within the total population studied, we observed a detection frequency of 69% HRV-A (n=44), 6% HRV-B (n=4) and 25% HRV-C (n=16). There was no difference in HRV species detection between the two patient groups ($p = 0.69$) (Figure 3.3). The large number of different types identified (43 HRV types in 64 patients) rendered any attempt to determine associations between specific HRV strains and disease presentations impossible.

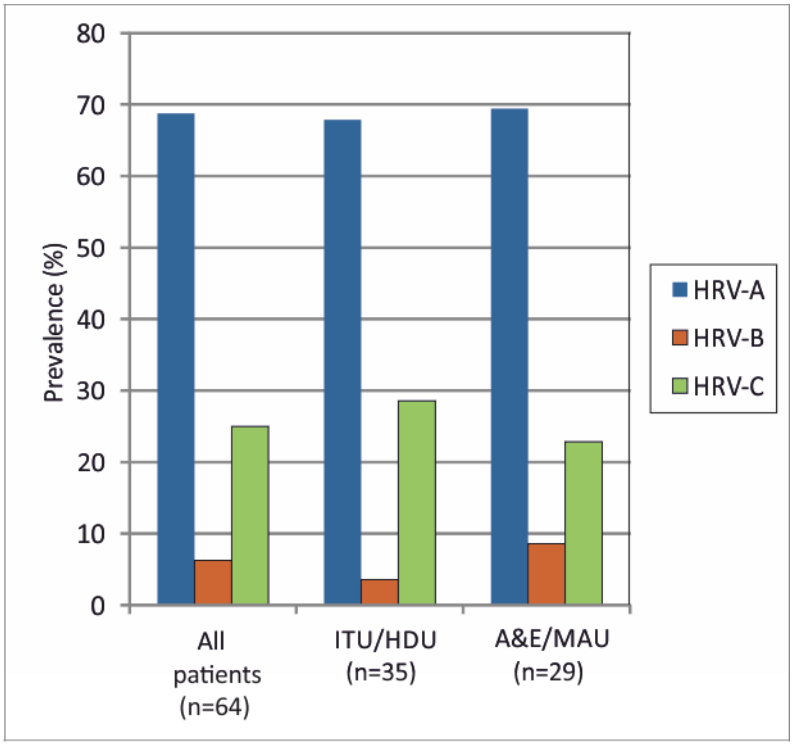


FIGURE 3.3: HRV species isolated from distinct patient groups. *Values are shown separately for ITU/HDU, A&E/MAU and total cohort separately.*

The vast majority of HRV positive samples were from those under the age of 5 (86%). The species prevalence within each age band followed the expected and previously observed pattern with HRV-A most prevalent and HRV-B the least. When grouped by age bands, patients did not segregate by admission to ITU or A&E.

30% of children sampled had evidence of a co-infection (n=19). Similar frequencies of co-infections were observed in patients admitted to ITU (74%) and patients seen in the acute care setting (67%). The most common co-infection observed in both settings was by AdV. This was not found to be correlated with an increased length of stay in hospital or more severe respiratory symptoms in this sample set. Notably, only two bacterial co-infections were identified; one with *Moraxella catarrhalis* and one with *Haemophilus influenzae.* Both of these occurred in children aged one to

two years admitted to ITU with severe respiratory distress requiring ventilatory support and both had an additional viral co-infection with AdV.

The Paediatric Index of Mortality (PIM2) scoring system is a tool for rating the severity of illness and mortality risk in children at the time of their admission to ITU (Slater et al., 2003). Consideration of a variety of clinical factors allows the calculation of a score correlating with a percentage likelihood of mortality. The likelihood of mortality as predicted by PIM2 score showed no significant difference between HRV species. Additionally, no relationship was evident between viral loads in respiratory samples and PIM2 score. However, these results are difficult to interpret due to a lack of detailed clinical information regarding the primary cause of illness in each child. For example, one child with an extremely high PIM2 score of 0.48317 was admitted to ITU with cardiac failure post cardiac arrest and a long-standing history of immunosuppression and mitochondrial cytopathy. Mitochondrial cytopathy predisposes to the development of LRTIs secondary to poor muscle tone and swallowing difficulties. However, the post-arrest status of this child and resulting cardiac failure are likely to have been the primary reasons for ITU admission and it is likely that the documented respiratory infection was a secondary concern which was already resolving (CT value of 38.64). Further studies would benefit from prospective collection of significantly more detailed clinical data.

3.3.3 Screening of CSF samples for HRV

As EV are an extremely common cause of aseptic meningitis, screening of CSF samples for HRV was undertaken in order to determine whether these closely related pathogens may also play a role in the development of this disease. A total of 1500 CSF samples from 2009 were screened for HRV in pools of 10. Of these, only one was positive, which originated from a female neonate (under the age of 3 months). The limited clinical data available identified a sepsis-like illness. It is not known whether the patient was immunocompromised. HRV was additionally present in a

throat swab from the same patient. However, stool samples and blood cultures were not obtained. These results indicate that HRV is an extremely rare and currently unrecognised potential cause of aseptic meningitis.

3.3.4 Isolation of HRV from faecal samples

Stool samples are referred to the SVC principally for diagnosis of suspected viral gastroenteritis or viral meningitis and also commonly in investigation of suspected infective outbreaks of diarrhoea and vomiting within the hospital setting.

A total of 26 from 288 stool samples screened were PCR-positive for HRV (9%). Positives were amplified and sequenced in the VP4/VP2 region for HRV species and type identification. All positive samples were isolated from hospitalised patients, of which over a third (35%) were obtained from patients under investigation for potential norovirus infections. The highest frequency of detection of HRV in an individual month was observed in September (10%). However, both May and December also displayed relatively high detection rates (7% and 8% respectively).

In the 26 positive samples, a total of 15 HRV VP4/VP2 types were identified (5 HRV-A, 2 -B and 8 -C). This suggests that rather akin to the pattern of circulation observed in respiratory specimens, a large number of types circulate concurrently each epidemic season (Savolainen et al., 2002). HRV species identified in the positive samples comprised 50% species C (n=13), 42% species A (n=11) and 8% (n=2) species B. Additionally, phylogenetic analysis reveals that HRV sequences obtained from stool samples were genetically diverse in the VP4/VP2 region (Figure 3.4).

Five HRV positive patients infected with highly similar strains of HRV-A1 were identified (marked with black dots on Figure 3.4). All of these patients presented in September 2010 with diarrhoea (Table 3.1) and had stool samples referred as part of the investigation of a norovirus outbreak. However, as the patients were treated in

three different locations, it is unlikely that these five samples represent an isolated outbreak of HRV infection.

In the absence of a control group with no gastrointestinal symptoms, it is not possible to conclusively link the presence of HRV in stool with gastrointestinal disease. However, 77% (n=20) had no documented evidence of concurrent respiratory tract infection, such as bronchiolitis or URTI. 27% (n=7) of HRV positive patients required treatment in an ITU setting. The majority of putative HRV gastrointestinal infections were observed in children under the age of 5 years (n=18, 69%), with a further peak observed in the over 65 age group (n=6, 23%). Nine of the HRV positive patients were patients who were initially referred as part of the investigation of suspected norovirus outbreaks and tested negative for norovirus.

Only three co-infections were noted in stool samples; two with AdV and one with *Salmonella enteritidis*. One child who was co-infected by AdV had symptoms of both an URTI and diarrhoea. Further screening performed by real-time PCR used for EV detection by Dr H. Harvala at the SVC detected 11 HRV positive stool samples from 333 tested (Harvala et al., 2012b). The false positivity for HRV in EV screening in the diagnostic setting arising from the high level of sequence conservation in the 5'UTR may have led to misrepresentation of the frequency of EV infections in clinical samples.

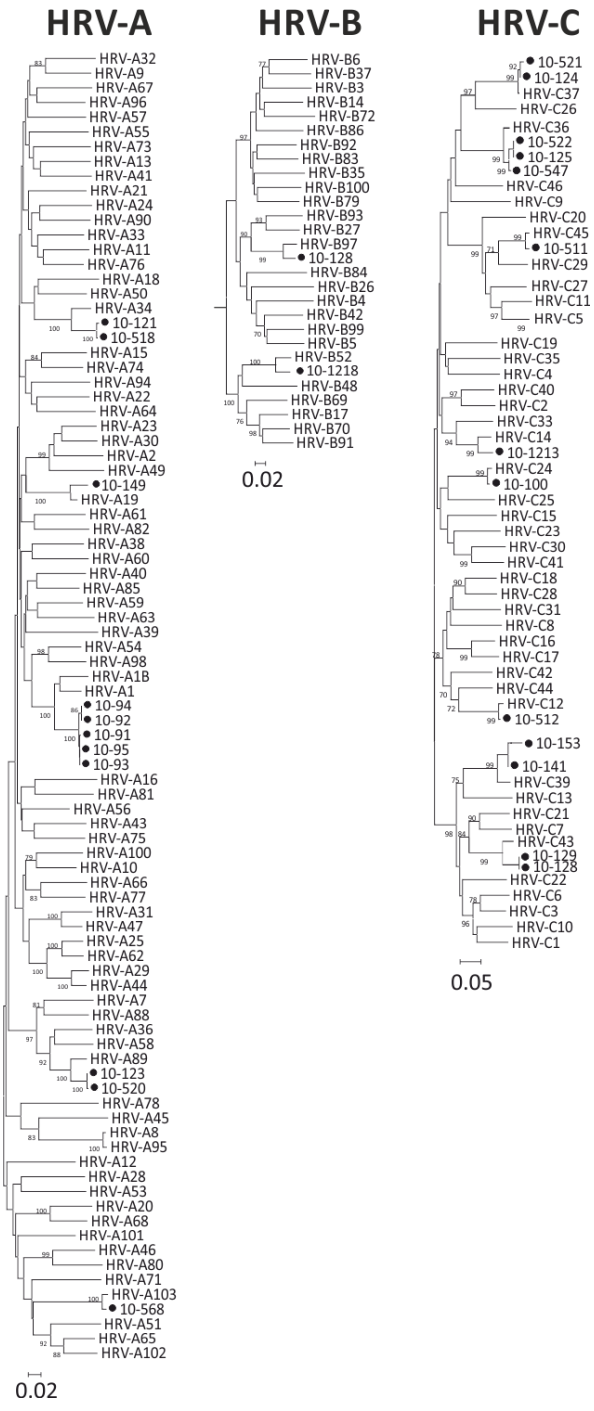


FIGURE 3.4: Neighbour joining phylogenetic trees of the VP4/VP2 region of HRV-A, HRV-B and HRV-C. *Sequences from isolates recovered from stool samples are marked with black dots. Trees were constructed using NC_001430 : EV70 as an outgroup. Branches are scaled by genetic distance.*

TABLE 3.1: Clinical characteristics of patients with HRV isolated from stool samples

Sample Number	Isolation month	Age	Clinical location	Clinical details
10_511	May 2010	6-10 years	RHSC	Diarrhoea
10_512	May 2010	7-12 months	RHSC	URTI, diarrhoea and vomiting
10_518	May 2010	3-5 years	RHSC	Diarrhoea and fever
10_520	May 2010	7-12 months	ESCITU	Diarrhoea and seizure
10_521	May 2010	<3 months	ESCITU	Diarrhoea, vomiting and fever
10_522	May 2010	<3 months	RHSC	Bronchiolitis and vomiting
10_547 [a]	May 2010	7-12 months	RHSC	Diarrhoea and vomiting
10_568 [a]	May 2010	3-5 years	RHSC	Diarrhoea and vomiting
10_91 [a]	Sept 2010	>65 years	RIE	Diarrhoea
10_92 [a]	Sept 2010	37-65 years	ITU	Diarrhoea and fever
10_93 [a]	Sept 2010	>65 years	ITU	Diarrhoea and fever
10_94	Sept 2010	1-2 years	RHSC	Diarrhoea and fever
10_95	Sept 2010	<3 months	RHSC	Diarrhoea and vomiting
10_100	Sept 2010	7-12 months	RHSC	Loose stools and irritability
10_128	Sept 2010	<3 months	RHSC	Bronchiolitis and loose stools
10_129[b]	Sept 2010	1-2 years	RHSC	Fever (post-transplant)
10_141[c]	Sept 2010	3-6 months	RHSC	Diarrhoea
10_149[b]	Sept 2010	1-2 years	RHSC	URTI and loose stool
10_153 [a]	Sept 2010	>65 years	ITU	Diarrhoea and fever
10_121	Dec 2010	3-6 months	ESCITU	Bronchiolitis and vomiting
10_123	Dec 2010	7-12 months	ESCITU	Bronchiolitis and vomiting
10_124	Dec 2010	7-12 months	RHSC	Gastroenteritis
10_125	Dec 2010	7-12 months	RHSC	Diarrhoea and seizure
10_128 [a]	Dec 2010	>65 years	RIE	Diarrhoea and vomiting
10_1213 [a]	Dec 2010	>65 years	RIE	Diarrhoea and vomiting
10_1218 [a]	Dec 2010	>65 years	RIE	Diarrhoea and vomiting

RIE – Royal Infirmary of Edinburgh, RHSC – Royal Hospital for Sick Children (Edinburgh), ITU – Intensive Therapy Unit (RIE), ESCITU – Edinburgh Sick Children ITU (RHSC)

[a] Nine patients were investigated as part of suspected norovirus outbreaks

[b] Two co-infections with Adenovirus were identified

[c] One co-infection with *Salmonella enteritidis* indentified

In order to ascertain whether HRV obtained from stool samples was present at concentrations consistent with the possibility of gastrointestinal infection, viral loads were recorded and compared for both HRV and EV positive samples. There was no significant difference in viral load (as measured by real-time PCR CT value) between

HRV (28.5; standard deviation ±3.5) and EV (28.9; standard deviation ±1.9) positive samples (p=0.77). However, one-step real-time PCR specific for the HRV 5'UTR performed as per SVC protocols detected only 42% (n=11) of the 26 HRV positive samples identified by nested RT-PCR. Despite the high frequency of HRV detection observed in all three months, more extensive sampling would be required to determine whether HRV RNA is found in stool samples throughout the year.

3.3.5 Discrepancy of screening results between real time PCR in diagnostic setting and VP4/VP2 screening PCR

In order to investigate the characteristics of concurrent HRV circulation in respiratory illnesses during September and December 2010, 731 respiratory samples were screened for HRV. A total of 572 of these had previously undergone screening procedures at the SVC using the standard HRV 5'UTR real-time RT-PCR. Of these, 115 (20%) were recorded as HRV positive and therefore typed by sequencing of the amplified VP4/VP2 region. Eight of these positive samples could not be amplified in VP4/VP2 and repeat screening with the 5'UTR nested RT-PCR previously described was negative (Wisdom et al., 2009b).

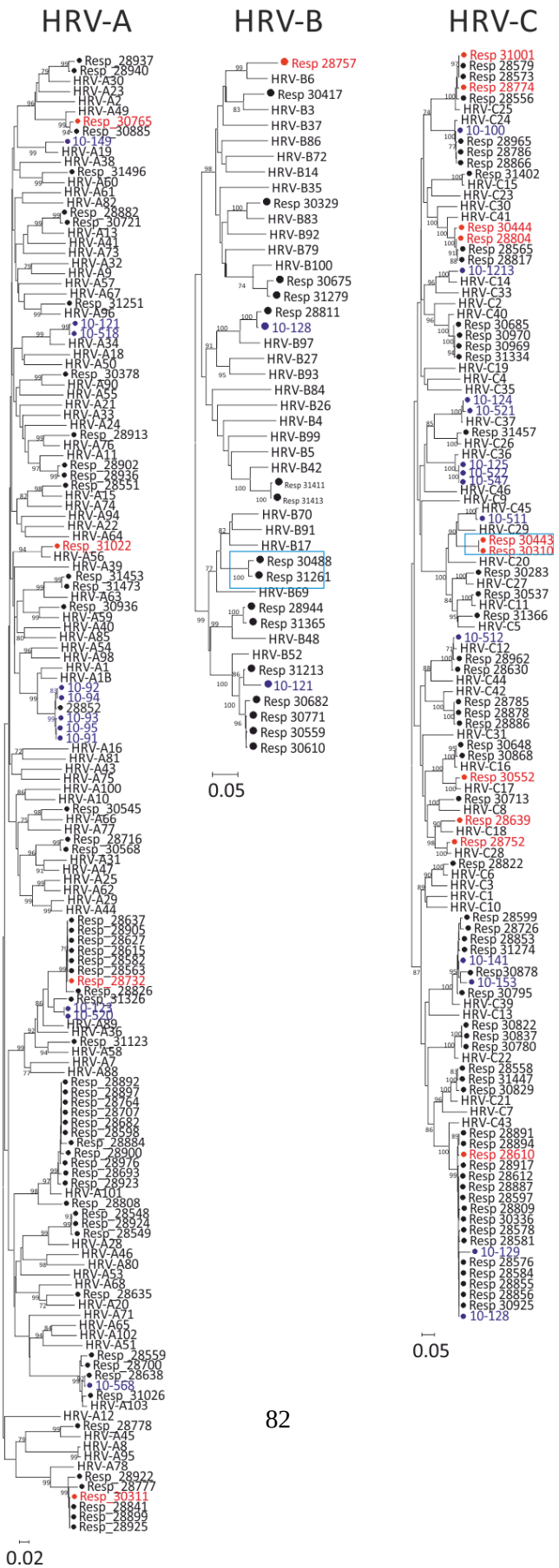
The 457 samples recorded as negative on routine diagnostic screening were combined into pools of 6 and screened by nested 5'UTR RT-PCR. 15 of these samples (3%) were actually HRV positive and typed in the VP4/VP2 region (shown in red on Figure 3.5). These 15 samples comprised 4 of species A, 10 species C and one species B.

The samples which were negative on initial 5'UTR PCR screening were genetically diverse, from all three HRV species and did not segregate in VP4/VP2. Within HRV-A, three out of the four previously negative samples grouped with samples that were identified by 5'UTR real-time PCR screening. Within HRV-C, five initially false-negative samples grouped with three types that had been identified by the

5'UTR real-time PCR screening. This suggests that the detection failure may be secondary to low viral load in these cases.

However, 6 false-negative HRV-C samples belonging to four types did not group with any previously identified HRV-C positive samples. All four of these belonged to groups that have been previously documented to be recombinant HRV-Ca types (Wisdom et al., 2009a; McIntyre et al., 2010, 2013b). Analysis of a small fragment of the 5'UTR that is additionally amplified during the VP4/VP2 typing PCR confirmed these 5'UTR groupings (data not shown). This may indicate that the addition of a single extra primer specific for HRV-Cc strains may not be sufficient and a redesign of primers used in this particular screening protocol may be warranted for use in routine HRV screening.

FIGURE 3.5: Neighbour joining phylogenetic tree showing the VP4/VP2 region of HRV positive samples from respiratory specimens and stool specimens. *NC_001430: EV70 was used as an outgroup for analysis for all trees and branch to tree root has been collapsed for ease of reference. Positive samples from the SVC respiratory archive are shown in bold black, positives from stool screening are shown in bold blue. HRV positive samples which were negative on initial real-time HRV 5'UTR RT-PCR screening are shown in red. Two HRV-B and two HRV-C VP4/VP2 sequences belonging to putative new types are outlined by a blue box (later designated HRV-Bpat4 and HRV-Cpat17 – see Chapter 4). Branches are scaled by genetic distance.*


HRV-A
HRV-B
HRV-C
Resp_28937
Resp_28940
HRV-A30
HRV-A23
HRV-A7
HRV-A49
Resp_30765
Resp_30885
10-149
HRV-A19
HRV-A38
Resp_31496
HRV-A60
HRV-A61
HRV-A82
Resp_28882
Resp_30721
HRV-A13
HRV-A41
HRV-A73
HRV-A32
HRV-A9
HRV-A57
HRV-A67
Resp_31251
HRV-A96
10-121
10-518
HRV-A34
HRV-A18
HRV-A50
Resp_30378
HRV-A90
HRV-A55
HRV-A21
HRV-A33
HRV-A24
Resp_28913
HRV-A76
HRV-A11
Resp_28902
Resp_28936
Resp_28551
HRV-A15
HRV-A74
HRV-A94
HRV-A22
HRV-A64
Resp_31022
HRV-A56
HRV-A39
Resp_31453
Resp_31473
HRV-A63
Resp_30936
HRV-A59
HRV-A40
HRV-A85
HRV-A54
HRV-A98
HRV-A1
HRV-A1B
10-92
10-94
28852
10-93
10-95
10-91
HRV-A16
HRV-A81
HRV-A43
HRV-A76
HRV-A100
HRV-A10
Resp_30545
HRV-A66
HRV-A77
Resp_28716
Resp_30568
HRV-A31
HRV-A47
HRV-A25
HRV-A62
HRV-A29
HRV-A44
Resp_28637
Resp_28905
Resp_28627
Resp_28615
Resp_28582
Resp_28563
Resp_28732
Resp_28826
Resp_31326
10-123
10-520
HRV-A89
HRV-A36
Resp_31123
HRV-A58
HRV-A7
HRV-A88
Resp_28892
Resp_28897
Resp_28764
Resp_28707
Resp_28682
Resp_28598
Resp_28884
Resp_28900
Resp_28976
Resp_28693
Resp_28923
HRV-A101
Resp_28808
Resp_28548
Resp_28924
Resp_28549
HRV-A28
HRV-A46
HRV-A80
HRV-A53
HRV-A68
Resp_28635
HRV-A20
HRV-A71
HRV-A65
HRV-A102
HRV-A51
Resp_28559
Resp_28700
Resp_28638
10-568
Resp_31026
HRV-A103
HRV-A12
Resp_28778
HRV-A45
HRV-A8
HRV-A95
HRV-A78
Resp_28922
Resp_28777
Resp_30311
Resp_28841
Resp_28899
Resp_28925
0.02
Resp 28757
HRV-B6
Resp 30417
HRV-B3
HRV-B37
HRV-B86
HRV-B72
HRV-B14
HRV-B35
Resp 30329
HRV-B83
HRV-B92
HRV-B79
HRV-B100
Resp 30675
Resp 31279
Resp 28811
10-128
HRV-B97
HRV-B27
HRV-B93
HRV-B84
HRV-B26
HRV-B4
HRV-B99
HRV-B5
HRV-B42
Resp 31411
Resp 31413
HRV-B70
HRV-B91
HRV-B17
Resp 30488
Resp 31261
HRV-B69
Resp 28944
Resp 31365
HRV-B48
HRV-B52
Resp 31213
10-121
Resp 30682
Resp 30771
Resp 30559
Resp 30610
0.05
Resp 31001
Resp 28579
Resp 28573
Resp 28774
Resp 28556
HRV-C25
HRV-C24
10-100
Resp 28965
Resp 28786
Resp 28866
Resp 31402
HRV-C15
HRV-C23
HRV-C30
HRV-C41
Resp 30444
Resp 28804
Resp 28565
Resp 28817
10-1213
HRV-C14
HRV-C33
HRV-C2
HRV-C40
Resp 30685
Resp 30970
Resp 30969
Resp 31334
HRV-C19
HRV-C4
HRV-C35
10-124
10-521
HRV-C37
Resp 31457
HRV-C26
HRV-C36
10-125
10-547
HRV-C46
HRV-C9
HRV-C45
10-511
HRV-C29
Resp 30443
Resp 30310
HRV-C20
Resp 30283
HRV-C27
Resp 30537
HRV-C11
Resp 31366
HRV-C5
10-512
HRV-C12
Resp 28962
Resp 28630
HRV-C44
HRV-C42
Resp 28785
Resp 28878
Resp 28886
HRV-C31
Resp 30648
Resp 30868
HRV-C16
Resp 30552
HRV-C17
Resp 30713
HRV-C8
Resp 28639
HRV-C18
Resp 28752
HRV-C28
Resp 28822
HRV-C6
HRV-C3
HRV-C1
HRV-C10
Resp 28599
Resp 28726
Resp 28853
Resp 31274
10-141
Resp30878
10-153
Resp 30795
HRV-C39
HRV-C13
Resp 30822
Resp 30837
Resp 30780
HRV-C22
Resp 28558
Resp 31447
Resp 30829
HRV-C21
HRV-C7
HRV-C43
Resp 28891
Resp 28894
Resp 28610
Resp 28917
Resp 28612
Resp 28887
Resp 28597
Resp 28809
Resp 30336
Resp 28578
Resp 28581
10-129
Resp 28576
Resp 28584
Resp 28855
Resp 28856
Resp 30925
10-128
0.05

3.3.6 Analysis of concurrent HRV circulation in respiratory samples during September and December 2010

A total of 142 HRV positive samples were identified in retrospective screening of 731 respiratory samples from September and December 2010. The overall detection frequency seen was relatively higher than that observed in stool samples (Table 3.2). However, the distribution of the three species of HRV was highly similar. HRV-C was the most frequently detected in both, closely followed by HRV-A, while HRV-B was relatively rare (Table 3.2). Additionally, in both sample types, HRV was relatively more common in September than December.

TABLE 3.2: Comparison of detection rates between HRV positives isolated from stool and respiratory samples

	Stool specimens	**Respiratory specimens**
HRV positive samples	9.0% (n=26)	19.4% (n=142)
% HRV-A	42.3% (n=11)	41.6% (n=59)
% HRV-B	7.7% (n=2)	12.0% (n=17)
% HRV-C	50.0% (n=13)	46.5% (n=66)
HRV positive samples – September	12.2% (n=11)	36.3% (n=81)
HRV positive samples – December	7.8% (n=7)	12.0% (n=61)

HRV positive samples observed represented 51 HRV types (21 HRV-A, 9 HRV-B and 21 HRV-C). Of the 13 HRV types detected in stool specimens from September and December 2010, only 6 were seen co-circulating in respiratory samples from the same month (marked on Figure 3.5). In addition, only two potential matches between a stool and respiratory sample could be identified. This was surprising, as although there was a low incidence of respiratory symptoms within the group of patients with HRV positive stool specimens, 7 of these were treated in ITU. The majority of ITU admissions with fever or other infective symptoms are routinely

subjected to a full septic screen comprising of respiratory, stool and urine samples with blood cultures. One HRV-A1 positive patient had the virus identified in both stool and upper respiratory secretions (sample numbers 10_94 and Resp_28852). In addition, HRV-C39 was isolated from both stool and upper respiratory secretions of a child aged 3-6 months admitted to the RHSC (sample numbers 10_141 and Resp_28726). Although matched respiratory samples could not be identified for HRV-C24 positive stool samples, three HRV-C24 positive respiratory samples were isolated from patients also admitted to the RHSC in the same month. HRV positive samples of the same type from the same month were also observed for HRV-A89, -B52 and –C39.

3.4 Discussion

3.4.1 Epidemiology and clinical correlations of HRV respiratory infections in Edinburgh

Throughout the course of this work, we have completed three studies of the epidemiology and clinical correlations of HRV respiratory infections in Edinburgh, including a study of the distribution of HRV species between ITU and A&E patients and investigations of concurrently circulating HRV in the same time period as HRV positive stool samples. A number of samples included in the study of HRV respiratory samples matched with stool samples had been previously screened for the presence of HRV by real-time RT-PCR specific for the 5'UTR of HRV. Of particular concern is the observation that this commonly used diagnostic PCR failed to detect a number of HRV positives. Indeed, around 3% of the samples that had been reported as negative by real-time PCR were actually HRV positive. Published literature on this subject indicates that this is not an isolated occurrence and not restricted to this particular PCR. Of 10 commonly used HRV 5'UTR primer pairs, no single pair could reliably detect all HRV present in a testing panel (Faux et al.,

2011)[13]. In accordance with our observations (Figure 3.5), the erroneous negative results did not correspond to a particular HRV species or type. Although the primer sets published previously (Wisdom et al., 2009b, 2009a) and used with modifications described within have not been independently tested for sensitivity and specificity, other work undertaken within our group found that the 5ˈUTR and VP4/VP2 nested PCRs were consistently able to detect RNA transcripts of all species of HRV in samples with close to 10^3copies/µL (McLeish et al., 2012).

Within the set of randomly selected respiratory samples from Edinburgh, HRV was found in 23% of cases, which falls within the range of 7.9% - 64.3% previously described (Alper et al., 2008; du Prel et al., 2009; Brittain-Long et al., 2010; Chan et al., 2012; Chidlow et al., 2012; Renwick et al., 2007; Lamson et al., 2006; Arden et al., 2006; Kistler et al., 2007a). We found that HRV was significantly more common in hospitalised patient populations than in the community based patients. However, the number of community based patients with referred respiratory samples was small and only four positive samples were identified in this group. Although the incidence of HRV is likely to be significantly higher in cases of mild URTI, patients with mild illness do not usually present and are not routinely sampled. Previous reports indicate that the incidence of HRV infection in asymptomatic individuals may vary from 2% to over 50% (Fry et al., 2011; Alper et al., 2008; Peltola et al., 2008; Brittain-Long et al., 2010; Chidlow et al., 2012). A valuable addition to future work would be the prospective collection of both samples and data from a community based control group, including both symptomatic and asymptomatic patients.

In accordance with previous studies, HRV-A was generally the most prevalent HRV species in respiratory samples. However, in both stool samples and respiratory samples from September 2010, HRV-C was the most commonly isolated species

[13] Primers in use in both the research and diagnostic setting in Edinburgh were not included in this published analysis.

(Figure 3.5). There was no difference in the distribution of infecting HRV species by age. Despite several published studies suggesting a correlation between HRV-C infection and more severe clinical illness, we found no significant difference between HRV species distribution or co-infection rate in patients admitted to ITU or A&E. Although ITU admission was presumed to correlate with illness severity, it may represent a somewhat crude estimate in some respects. For instance, some patients are likely to have required ITU care for another, unrelated condition, with the documented respiratory infection being a secondary concern. The limited amount of clinical data that was able to be collected for each patient retrospectively limited the analysis. Future studies would benefit from prospective data collection.

The results of these studies have several notable differences from a similar study of 456 Edinburgh respiratory samples carried out in 2006/2007 (Wisdom et al., 2009b). Although this study was carried out less than 5 years prior to the studies within this thesis, they found that HRV-A were considerably more frequent among patients admitted to ITU and HRV-C strains were entirely absent. In addition, they found a significantly elevated incidence of HRV infection in the over 65 age group, whereas we observed HRV in only 4% of this group. However, certain other features, such as a lack of male or female bias and the generally more frequent isolation of HRV-A are consistent not only between these two studies, but are widely repeated observations. In accordance with our study, Wisdom and co-workers also identified a large number of genetically distinct co-circulating HRV strains (Wisdom et al., 2009b). Determining the finer details of HRV circulation patterns, epidemiological features and clinical correlates are unlikely to be achieved with small, isolated clinical studies. An unprecedented volume of data would be required in order to effectively investigate the correlation of specific HRV strains with specific disease presentations. A large, multicentre, prospective epidemiological study carried out over a period of several years and over many geographical locations, using standardised PCR reagents and primers would potentially allow characterisation of

such features. However, such an endeavour, whilst undoubtedly worthwhile would likely prove to be both logistically challenging and labour intensive.

3.4.2 Screening for HRV in cerebrospinal fluid samples

Despite the detection of HRV RNA in stool samples from patients with acute flaccid paralysis and *in vitro* propagation of live virus from sewage under conditions optimised for the growth of EV (Victoria et al., 2009; Blomqvist et al., 2009), there was only very little evidence of HRV in CNS infection. Indeed, the one positive sample that was recovered led to an overall detection frequency of HRV in aseptic meningitis of less than 0.5%, which although low is still remarkable. The identification of only one positive sample led to the consideration of the possibility of contamination or laboratory error. However, the individual sample underwent the process of RNA extraction, nested PCR and sequencing on two independent occasions and produced an identical result. In addition, no other laboratory work with HRV positive samples was being carried out during this period. Overall, these results suggest that although it is significantly more rarely observed than EV, HRV is a potential cause of neonatal sepsis-like illness and potentially other CNS disease.

3.4.3 Isolation of HRV from faecal specimens

Since the publication of our work on this subject (Harvala et al., 2012b), two further studies detailing the presence of HRV in diagnostic stool specimens have been published (Lau et al., 2012; Honkanen et al., 2013). In addition, one analysis of HRV positive children in Malaysia contained an incidental finding that 34.8% had a primary symptomatology of diarrhoea and vomiting, rather than respiratory symptoms (Chan et al., 2012). Although the high incidence of nearly 10% we observed was surprising, the possibility that HRV may be routinely present in stool samples has been evident for a long time. Not only has HRV been isolated from sewage samples (Blomqvist et al., 2009), from cases of acute flaccid paralysis (Victoria et al., 2009) and from stool samples from critically ill children (Tapparel et

al., 2009c; Broberg et al., 2011), but information available from the ATCC archive indicates that the original HRV-33 prototype strain was isolated from a pool of human stools collected in Maryland in 1939 (strain VR-330). This suggests that HRV has been potentially circulating in faecal samples largely undetected for many years.

HRV-C was the most frequently isolated HRV species from stool samples and accounted for 50% of HRV positive samples. Lau et al also found that HRV-C was detected most frequently in stool samples (Lau et al., 2012). However, in a large study of 4184 stool samples isolated from Finnish children, HRV-A was the most frequently detected HRV species (Honkanen et al., 2013). We found no specific species or strain to be associated with gastrointestinal infection. Although Honkanen and co-workers reported that HRV-A61 was the most commonly amplified HRV type (Honkanen et al., 2013), we did not observe any occurrences of this type within our sample set. The genetic diversity of HRV strains in stool suggests that there may not be specifically enteropathic HRV strains and indeed, all HRV may be capable of infecting the gastrointestinal tract under favourable circumstances. Although there was very little overlap between circulating infections with identical HRV strains in stool and respiratory samples between September and December, the same broad trends were observed (Table 3.2). Not only was HRV-C over-represented in both sample sets but the detection frequency of HRV was uniformly higher in September. This correlates with a previously observed high frequency of HRV infection in Edinburgh in September (Wisdom et al., 2009b), which is possibly linked to the declining temperatures and the start of the school year leading to increased person to person contact in a warm, humid environment. The relatively high incidence of HRV-C in both sample sets may be an incidental finding related to small sample sizes or it may be indicative of a genuinely higher circulation of HRV-C near the close of 2010.

HRV are traditionally considered to be acid labile and indeed, retention of infectivity at pH of less than 5 was often used to differentiate EV from HRV in clinical diagnostics. Previous findings that may have raised the possibility of HRV as a gastrointestinal pathogen have been attributed merely to swallowed respiratory secretions or discarded tissues (Blomqvist et al., 2009). However, our observation of infrequent reporting of respiratory symptoms in cases of HRV positive stool suggests that swallowed respiratory secretions may not play a significant role. This assertion is further strengthened by the observation of high viral loads, which are incompatible with swallowed HRV being diluted by liquid gut contents. In addition, if simply suffering an upper respiratory HRV infection could lead to the routine isolation of HRV from stool or sewage samples, the detection frequency of HRV in stool samples would surely be substantially higher. Neither our study, nor any published since can claim to prove causality by observation of this association with HRV positive stool samples and gastrointestinal pathology. However, there was no statistically significant difference between viral loads in HRV and EV positive samples (Harvala et al., 2012b). As EV is known to replicate in the gastrointestinal tract, this observation lends weight to the assertion that HRV may be a causative factor in these infections and should potentially be further investigated.

The isolation of HRV from stool samples of symptomatic patients at a similar rate to the isolation of EV indicates the potential value of addition of HRV to the diagnostic screening procedures currently employed. Some diagnostic laboratories use the presence of EV in stool samples as an adjunct to screening CSF samples in the diagnosis of aseptic meningitis. As HRV appears to potentially have a significant role in gastrointestinal infection, this approach may lead to overdiagnosis of EV meningitis in cases where the sample is not further typed. Indeed, 15% of reported EVs identified by diagnostic screening PCR carried out at the SVC were subsequently found to be HRV (Harvala et al., 2012b).

In conclusion, in common with previously published data, studies of HRV epidemiology in respiratory infections did not reveal any association of a particular strain or species with illness severity or clinical presentation. Significantly larger studies which span a wide geographical area over the course of many years may be the only possible method of determining whether these patterns actually exist and if so, what their nature may be. In contrast to the high incidence of EV observed in cases of CNS infection, HRV was observed at an extremely low frequency in CSF, suggesting that it may be a rare cause of CNS disease. Intriguingly, we found evidence of HRV in around 9% of stool samples from patients with symptomatic gastrointestinal infections and viral loads were similar to that observed for EV infections. Although this was the first published report of the incidence of HRV RNA in the GI tract, two other studies published subsequently have reported similar findings (Lau et al., 2012; Honkanen et al., 2013). Further studies of HRV infection in stool may focus on larger sample sets to determine seasonality and potential circulation patterns, as well as clinical correlates. Investigations in the biological differences in acid stability between these strains and those commonly isolated from respiratory infections may allow determination of the genetic factors that make such infections possible.

Chapter 4
Proposals for the Classification of All Three Species of Human Rhinovirus into Genotypically Assigned Types

4.1 Introduction

4.1.1 The standard approach to typing in HRV

By the 1960's, a large number of antigenically distinct HRV isolates had been identified and development of a system for their classification became essential. On the basis of neutralisation properties, HRV were divided into 100 serotypes by 1987 (Kapikian et al., 1967; Hamparian et al., 1987) and prototype strains of each were lodged at the American Tissue Culture Collection (ATCC).

For many years, serological methods were the standard technique used for typing HRV and EV isolates. Serological typing comprises isolation of the virus in cell culture, followed by neutralisation with intersecting pools of type-specific antisera and confirmation with a single type-specific antiserum. This approach was extremely time-consuming, laborious and fraught with difficulty. HRV are challenging to propagate in cell-culture, requiring specific HRV susceptible cell-lines and certain growth requirements (for example, an incubator at 33°C and roller-bottle culture vessels). With over 100 strains recognised, accurately typing even one isolate was a formidable task and standardised sera are in limited supply. The result was that typing was not generally undertaken in HRV and only rarely in EV.

When undertaken, serotyping in EV took 1-2 weeks. In most cases, an umbrella diagnosis of EV infection was deemed sufficient as the results of serotyping had a negligible impact on patient management. The most clinically useful application was

to distinguish between PV and non-polio enterovirus (NPEV) infections in enterovirus surveillance.

Serotyping also routinely encountered numerous problems leading to reports of "untypeable" isolates. Failure due to virus aggregation within the sample was commonly encountered (Wallis and Melnick, 1967). This could be overcome by filtration or treatment with sodium deoxycholate (JG Kapsenberg, A Ras, 1980). Similarly, failure arising from the presence of a mixture of viruses within the sample could be circumvented by plaque purification or limiting dilutions (Blomqvist et al., 2008). However, both these solutions created additional time-consuming steps in an already lengthy and complex procedure.

As a further difficulty, antisera that were originally developed against strains of HRV and EV isolated in the 1950s and 1960s, increasingly showed reduced or non-reciprocal cross-neutralisation with contemporary isolates through antigenic drift. Several of these antigenic variants (later identified as belonging to a specific type by VP1 sequence analysis (Oberste et al., 2000)) were known as prime strains. These contemporary isolates cannot be neutralised by antisera to the prototype strain while antisera to the contemporary isolate can neutralise the prototype. This has been reported on several occasions for EV (Abraham and Holmes, 1974; Schmidt et al., 1966; Norder et al., 2002), and for HRV (Halfpap and Cooney, 1983). These "inter-type" strains showing intermediate serological properties likely represent variants in the process of diverging into antigenically different types.

Cell culture based screening routinely undertaken for EV and HRV relied on visual inspection of cultures for cytopathic effect and was therefore subjective, especially when used on a large scale. Serological based techniques also suffer from the same subjectivity, even in experienced hands. In one published survey of twelve labs, none reported the correct or expected serotyping results for all ten EV samples

included and an overall rate of accurate reporting of around 60% was recorded (van Loon et al., 1999). In addition, the requirement for multiple cell lines to screen for HRV and HEV simultaneously (as would be preferable in screening of stool samples), renders cell culture screening nearly impossible. PCR based screening methods offer an easier and more objective alternative.

Other alternatives proposed to serotyping include viral protein fingerprinting (Holland et al., 1998) and immunofluorescence (IFA) based methods (Kiang et al., 2009), both of which can be highly accurate. However, viral protein fingerprinting requires specialised instruments, generates radioactive waste and relies on comparison with available protein sequences, which are significantly less abundant than nucleotide sequence data. IFA based methods suffer from some of the same problems as serological typing, in that reagents are limited and not all known serotypes are able to be tested.

As PCR is a ubiquitous technique in most labs and online databases contain a wealth of nucleotide sequence data, a classification system based on sequence data was ideal. Such a system was developed in 1999 for EV (Oberste et al., 1999b).

4.1.2 Definition of new EV types based on sequence divergence in the VP1 region

In order to classify EV by sequence data, it was essential to consider the likely molecular determinants of serotype. Sequence of the VP1 region correlates with serotype designation in PV isolates (Kilpatrick et al., 1998) and other NPEV (Oberste et al., 1999b, 1999a; Kiang et al., 2009). Studies of sequence divergence in VP1 led to the proposal of a threshold of 25% nucleotide and 12% amino acid divergence for the identification of EV types (Oberste et al., 1999b). These thresholds are now routinely used and a great number of new EV types have been classified on this basis (Oberste et al., 2002, 2007; Brown et al., 2009; Oberste et al.,

2004c; Smura et al., 2007b). As some EV isolates have documented evidence of recombination within the capsid region (Bouslama et al., 2007; Blomqvist, 2003; Zhang et al., 2010), only VP1 is recommended to be used for typing in EV. Previous attempts to classify EV by VP2 sequences (Arola et al., 1996; Oberste et al., 1998) have been unsuccessful.

The classification of EV isolates by molecular methods has now replaced neutralization assays in standard practice. These methods have been found to consistently outperform serotyping, in terms of accuracy, speed and ability to classify new types (Oberste et al., 2000; Kiang et al., 2009). The method also has benefits in the arena of polio eradication, allowing investigators to distinguish between vaccine derived and wild PV and to track emerging recombinant forms. .

4.1.3 The requirement for genotypic classification of HRV

With such a broad scope of genetic diversity and wide range of distinct clinical manifestations, it is no longer practical or desirable to consider all HRV types as one biological and clinical entity. There are presently no known specific associations of any HRV types with a particular disease. However, as the range of serious clinical illness that can be attributed to HRV infection is becoming more fully appreciated, there will likely be a requirement to routinely screen for HRV in the diagnostic setting and to commence large-scale epidemiological studies to uncover circulation patterns and strain associations. An easy and practical system of classifying HRV into types, analogous to the system now in use for EV (Oberste et al., 1999a) will allow investigations of potential outbreaks and nosocomial transmission and investigation of type-specific biological properties such as identification of specific types with potentially increased virulence.

Methods which rely on cell culture isolation are not practical for detection and characterisation of HRV, especially HRV-C. Indeed, HRV-C cannot be isolated by

most standard cell culture techniques and propagation currently relies on sinus mucosal organ culture (Bochkov et al., 2011). *In vitro* propagation of HRV-C has previously been attempted in WI-38, WisL, BEAS-2B, A549, HeLa, MRC-5, primary human lung fibroblasts and bronchial, sinus and adenoidal epithelial cells with no success (Bochkov et al., 2011; Kistler et al., 2007a; McErlean et al., 2007). In other viruses, such as hepatitis C virus (HCV), which are challenging or impossible to grow in standard cell culture, genotyping has provided a useful alternative means of classification.

The studies presented within this chapter aimed to develop a system of classification informed by nucleotide divergence for all three species of HRV. In 2010, we published proposals for the classification of HRV-C into genotypically defined types (Simmonds et al., 2010). These have now been adopted by the Picornavirus Study Group as the primary mechanism for defining HRV-C types. In addition, we have extended this classification system to include HRV-A and HRV-B, leading to a full re-evaluation of type designations within the two species (McIntyre et al., 2013a). Since the original division of HRV into 100 serotypes, an additional 62 types have now been designated. These include 52 HRV types determined from the results of studies contained within this thesis. The term "serotype" implies identification and classification by direct investigation of antigenic properties. Therefore, we have simply used the term "type" throughout to represent HRV types which have been identified and classified by sequence data alone, as is common practice for newly identified EV types.

4.2 Materials and Methods

4.2.1 Selection of samples

A selection of both published and unpublished studies of HRV epidemiology and evolution in Edinburgh yielded a wealth of HRV VP4/VP2 sequences (Gaunt et al.,

2011; Wisdom et al., 2009b, 2009a; Harvala et al., 2012b; McIntyre et al., 2010). VP4/VP2 sequences from our studies (431 HRV-A and 113 HRV-B) were divided into groups based on phylogenetic clustering. Any isolate which did not show a close grouping with a prototype full genome was selected for amplification of the VP1 region.

In addition, the 508 available HRV-C VP4/VP2 sequences obtained from isolates within the SVC archive were divided into groups on the basis of phylogenetic clustering. Representatives from each cluster, presumed to be analogous to HRV-C types, were selected for VP1 amplification.

4.2.2 Amplification of the VP1 region

The full VP1 region of HRV-A and -B isolates was amplified and sequenced in two overlapping sequence fragments (as described in Chapter 2). For HRV-A28 samples, amplification of the 5' fragment was not possible with the standard protocol and a specific inner anti-sense primer was designed. In all cases, where amplification with GoTaq polymerase (Promega, UK) was unsuccessful, it was attempted using a SuperScript III (Invitrogen, UK) protocol.

Two differing protocols were used for the amplification of the VP1 region of HRV-C during the course of this work. The first HRV-C VP1 PCR used the standard protocols described above and primers labelled as C VP1 PCR 1 (Appendix 1: Primer sequences). This resulted in two sequence fragments with a 105 base overlap (McIntyre et al., 2010). As this assay was only successful in amplifying 72 of 89 samples, a second outer anti-sense primer was developed (labelled as C VP1 PCR 2 in Appendix 1: Primer sequences). This improved assay allowed the amplification of the full HRV-C VP1 region and only failed to amplify one HRV-C isolate tested. The second round of this nested PCR reaction was carried out with an annealing temperature of 45°C and in a 50 µl reaction volume. PCR products were then

separated at 150V for 45 minutes on a 2% agarose gel and DNA bands of 1000 bases were excised manually under UV transillumination. Excised DNA bands were purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, UK) and sequenced as previously described.

4.2.3 Sequence alignment, calculation of pairwise nucleotide p-distances and phylogenetic analysis

All available HRV VP4/VP2 and VP1 sequences were downloaded from Genbank in November 2012. The genome regions analysed in this study included the full VP1 region and a fragment of VP4/VP2 that is commonly used in studies of HRV epidemiology (Savolainen-Kopra et al., 2009a; Kaida et al., 2011; Piralla et al., 2011; Xiang et al., 2010a; Han et al., 2009; Mizuta et al., 2010; Henquell et al., 2012; Miller et al., 2009a; Blomqvist et al., 2009). Any sequences which were less than 90% complete across these regions (Table 4.1) or classified as non-functional were excluded.

TABLE 4.1 : ***Co-ordinates of capsid coding genome regions used in analysis and corresponding HRV reference strain***

HRV Species	Numbered by:	VP4/VP2 Region	VP1 Region
HRV-A	FJ445111 (HRV-A1)	627 – 1015	2337 – 3197
HRV-B	X01087 (HRV-B14)	625 – 1013	2296 – 3192
HRV-C	EF582385 (HRV-C4)	616 – 1004	2305 – 3125

All sequences were aligned in SSE v1.0 and pairwise nucleotide p-distances were calculated using the program SequenceDist within SSE v1.0 software package. Phylogenetic trees were constructed as previously described and all phylogenetic analysis was undertaken using NC_001430 (EV-D70) as an outgroup.

4.2.4 Analysis of recombination within the capsid region of HRV-A, -B and –C

Phylogenetic trees were visually inspected for evidence of incongruence between VP4/VP2 and VP1. Datasets containing these two non-consecutive regions were concatenated to give one continuous sequence, which was then additionally analysed with RDP and GARD software packages.

4.3 Results

4.3.1 Assigning nucleotide divergence thresholds in the VP1 region of HRV-A, -B and –C

In order to determine whether a distinct threshold that divided pairwise p-distance comparisons into intra- and inter-type values existed within each HRV species, distributions of pairwise nucleotide p-distances were constructed from a total of 435 HRV-A, 133 HRV-B and 206 HRV-C sequences (Figure 4.1). These showed a maximum within-species nucleotide p-distance of 39.4%, 33.9% and 42.9% respectively and minimum values between 10% and 14% divergence. Detailed inspection allowed identification of minimum values in the distribution and assignment of a nucleotide divergence threshold (marked in grey on Figure 4.1 A2, B2, C2). A threshold of 13% was identified for both HRV-A and HRV-C sequences, while the less divergent HRV-B isolates showed a minimum value that supported a 12% threshold. Although analysis of nucleotide divergence thresholds for HRV-C sequences was carried out in 2010 (Simmonds et al., 2010), the results presented herein represent a repeated analysis incorporating an additional 120 HRV-C VP1 sequences. However, despite larger scale of the renewed investigation, the observed nucleotide divergence threshold of 13% remains unchanged from our original proposals (Simmonds et al., 2010).

As the majority of the sequences included originated from a large variety of clinical and evolutionary studies and types show different prevalences, there was an unequal representation of different putative HRV types. The number of strains available for a specific type ranged from 1 to a maximum of 13 for HRV-A (A49), 10 for HRV-B (B69) and 22 for HRV-C (C3). Therefore, an additional analysis was carried out including only the prototype strains of all three species (coloured grey on Figure 4.2 A1, B1 , C1). Prototype strains of HRV-A and -B were the earliest isolated full genome sequence of a known type and in HRV-C, prototype strains were those defined in our original classification scheme (Simmonds et al., 2010). The distributions of pairwise nucleotide p-distances of prototype strains closely reproduced the larger analyses (Figure 4.1).

Distributions of pairwise nucleotide p-distances within the VP1 region of HRV-A and HRV-C sequences showed a single large peak representing inter-type comparisons. In contrast, HRV-B sequences had two large peaks within the inter-type distance range. Further analysis revealed that the smaller peak, between p-distance values of 0.190 and 0.260 corresponds to comparisons between HRV-B types within each of the four basally branching phylogenetic clusters (marked by grey boxes on Figure 4.3). The larger peak corresponds to comparisons between variants in different clusters.

When sequences from all three species of HRV were compared, the pairwise nucleotide p-distances fell into three distinct distributions (Figure 4.2). As expected, the lowest values represented comparisons from within the same type and the highest (between 43% and 56% divergence) were between isolates of different species. The large number of comparisons between 15% and 43% divergence represent isolates of the same species but differing type. The existence of a discrete inter-species p-distance threshold reinforces the notion that HRV species may be defined on the basis of nucleotide divergence in the capsid region.

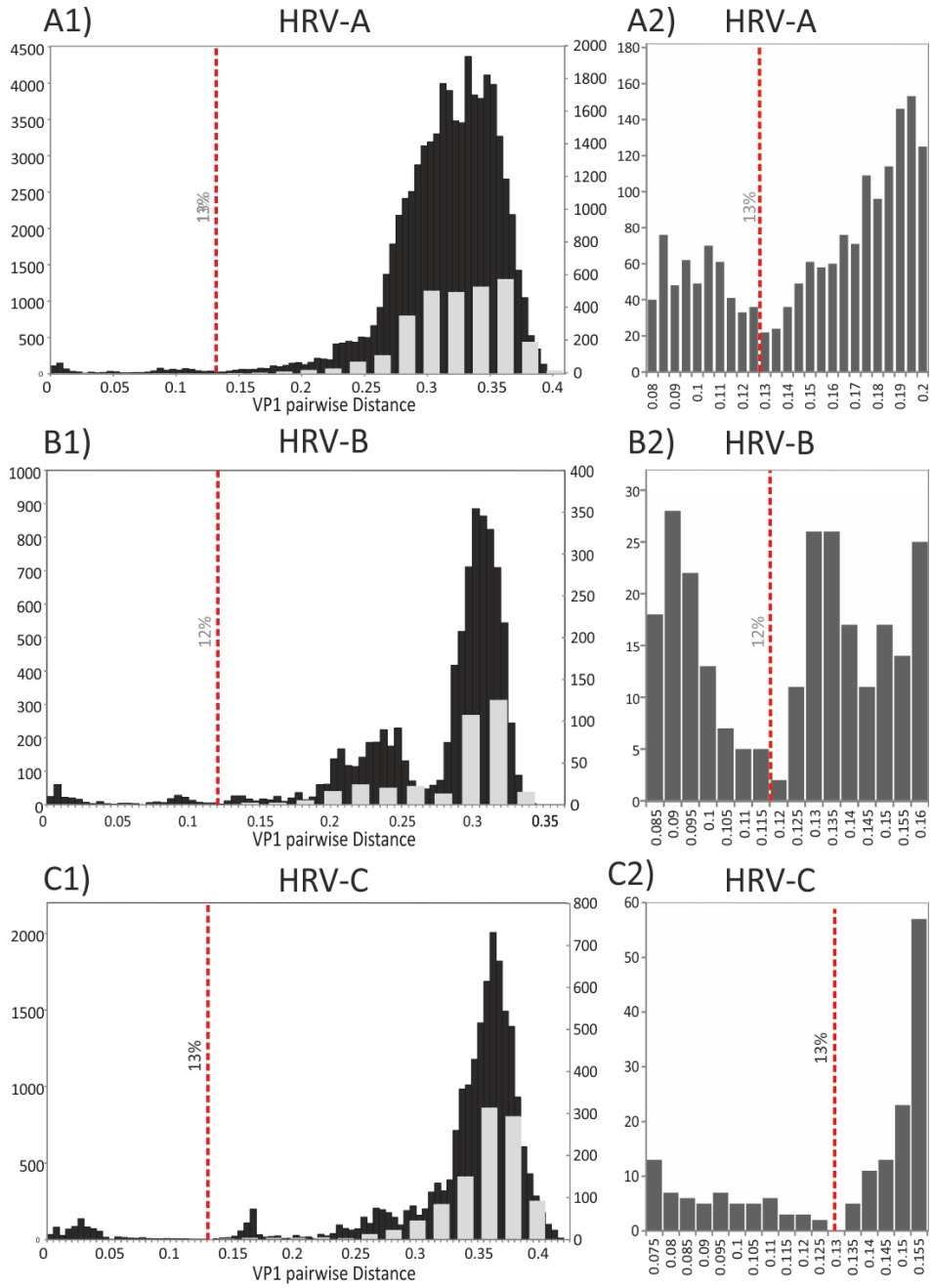


FIGURE 4.1: Distributions of pairwise nucleotide p-distances for the VP1 region of all three species of HRV. *HRV-A is shown in section A1, HRV-B in B1 and HRV-C in C1. Analysis including all available VP1 sequences is shown in black with y-axis scale on the left. Analysis including prototype strains only is shown in grey with y-axis scale on the right. A detailed analysis of the distribution values adjacent to the proposed thresholds for HRV-A (A2), HRV-B (B2) and HRV-C (C2) is also shown and the proposed nucleotide divergence threshold is shown by a red dotted line.*

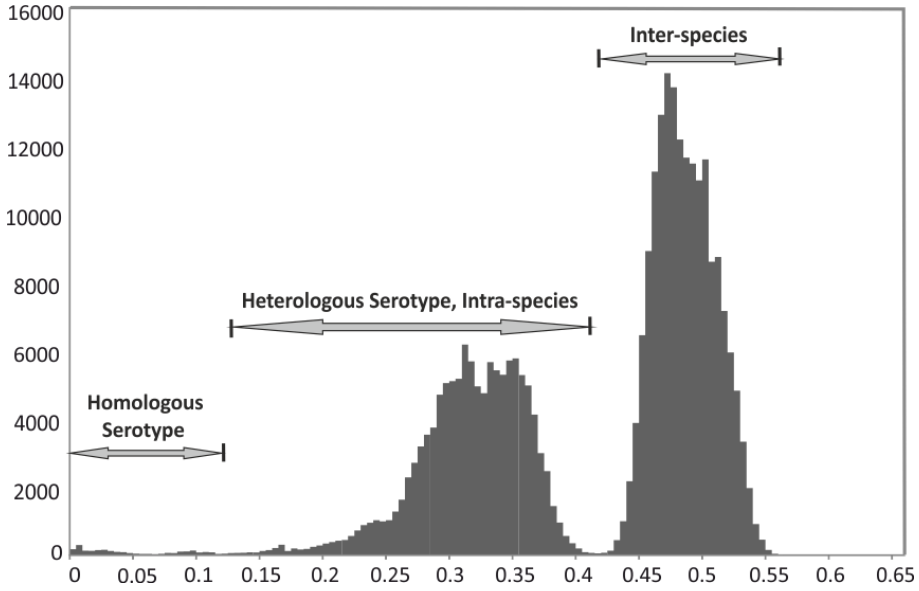


FIGURE 4.2: ***Distribution of pairwise nucleotide p-distances for the VP1 region of all HRV sequences.***

4.3.2 Phylogenetic analysis of the VP1 region

Phylogenetic trees were constructed of the VP1 region of HRV-A, -B and –C (Figure 4.3). All HRV sequences were divided into groups which were defined by bootstrap supported phylogenetic clades that closely matched types assigned by sequence distances. These groups were numbered either by clustering with a known prototype strain (HRV-A/B 1-100) or by order of submission of full genome (HRV-A101-106, HRV-B100-104 and HRV-C1-11) or VP1 sequence (HRV-C12-51). The majority of sequences which grouped together within the same putative type group showed intra-clade VP1 divergence of less than the proposed thresholds of 13% in HRV-A and HRV-C and 12% in HRV-B (Table 4.2). Most of these also showed a minimum inter-clade VP1 divergence (with the nearest neighbour type group) of greater than the proposed threshold. However, a number of HRV type groups (13, 2 and 1 in HRV-A, -B and –C respectively) violated the proposed VP1 thresholds (marked in red on Table 4.2 and further discussed in section 4.4.4).

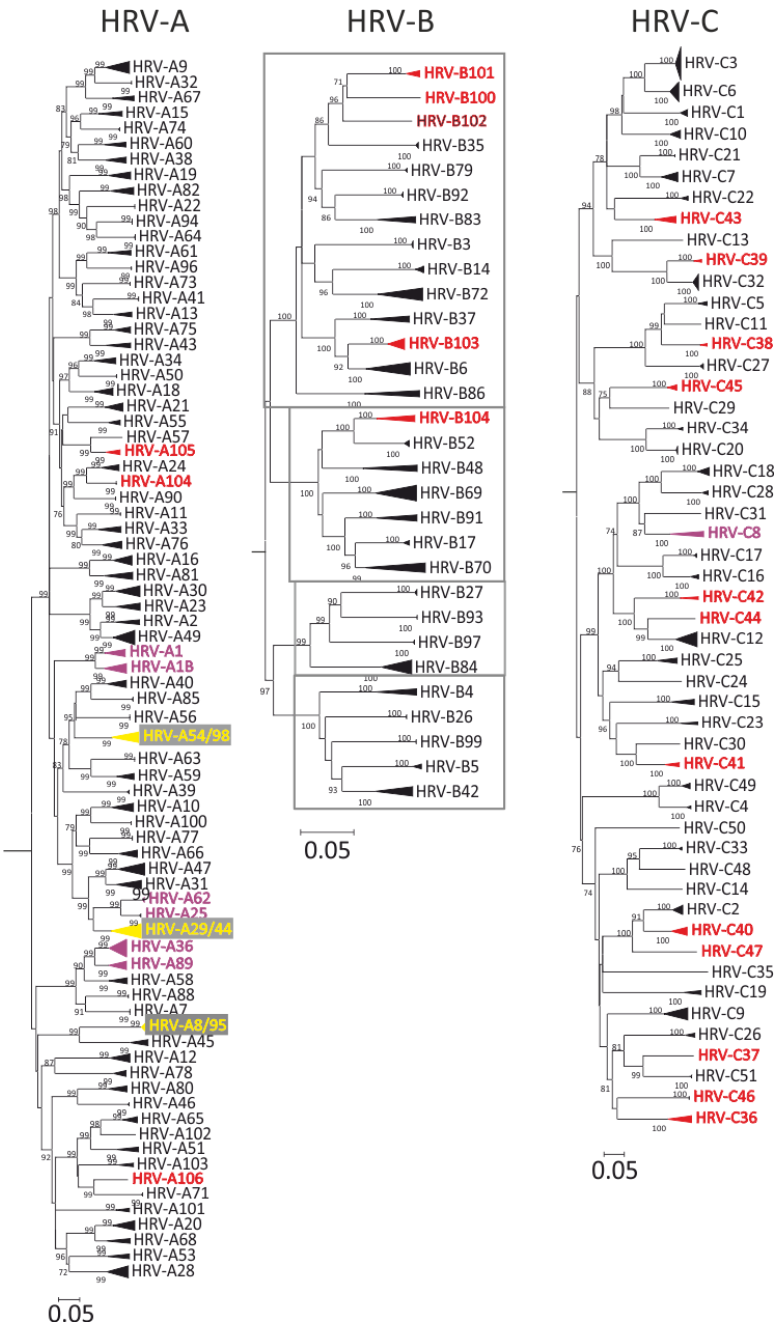


FIGURE 4.3: ***Phylogenetic trees showing the VP1 region of HRV-A, -B and –C.*** *Putative type groups and branch to tree root has been collapsed for ease of reference. Within HRV-B, phylogenetic clusters with pairwise p-distances ranging from 0.190 to 0.260 are bounded by grey boxes. Instances where two types have been combined to form a single type are marked in yellow (see Section 4.3.3). New HRV types defined on the basis of sequence divergence in VP1 are marked in red and HRV types with intermediate VP1 divergence are marked in purple. Branches are scaled by genetic distance.*

TABLE 4.2: Limits of intra- and inter-clade VP1 p-distance for HRV-A, -B and -C

HRV Type[a]	Max. intra-clade distance[b]	Min. inter-clade distance[c]
A1	**0.091**	**0.119**
A1B	**0.109**	**0.119**
A2	0.106	0.141
A7	0.005	0.182
A8	**0.004**	**0.015**
A9	0.109	0.148
A10	0.099	0.170
A11	0	0.172
A12	0.099	0.263
A13	0.082	0.171
A15	0.103	0.163
A16	0.089	0.172
A18	0.086	0.175
A19	0.098	0.227
A20	0.099	0.150
A21	0.095	0.205
A22	0.113	0.190
A23	0.107	0.143
A24	0.076	0.131
A25	**0.006**	**0.093**
A28	0.112	0.228
A29[d]	**0.007**	**0.072**
A30	0.111	0.143
A31	0.117	0.140
A32	0.001	0.148
A33	0.103	0.161
A34	0.096	0.158
A36	**0.115**	**0.120**
A38	0.094	0.192
A39	0.009	0.234
A40	0.121	0.159
A41	0.006	0.171
A43	0.110	0.161
A67	0.095	0.178
A68	0.104	0.150
A71	0.005	0.172
A73	0.003	0.182
A74	0.006	0.164
A75	0.079	0.161
A76	0.092	0.161
A77	0.001	0.185
A78	0.116	0.263
A80	0.087	0.203
A81	0.129	0.172
A82	0.105	0.195
A85	0.006	0.159
A88	0.001	0.182
A89	**0.088**	**0.120**
A90	0	0.174
A94	0	0.158
A95	**0**	**0.015**
A96	0	0.180
A98	**0**	**0.107**
A100	0	0.147
A101	0.061	0.261
A102	0	0.151
A103	0.069	0.216
A104	0	0.131
A105	**0.067**	**0.129**
A106	0	0.172
B3	0.005	0.218
B4	0.094	0.196
B5	0.014	0.166
B6	**0.088**	**0.113**
B14	0.018	0.186
B17	0.008	0.124
B102	0	0.163
B103	**0.035**	**0.113**
B104	0.085	0.121
C1	0.052	0.229
C2	0.057	0.165
C3	0.039	0.146
C4	0.016	0.139
C5	0.042	0.170
C6	0.066	0.146
C7	0.075	0.152
C8	**0.132**	**0.219**
C9	0.110	0.282
C10	0.037	0.205
C11	0	0.170
C12	0.101	0.197
C13	0.005	0.264
C14	0	0.215
C15	0.092	0.271
C16	0.044	0.147
C17	0.018	0.147
C18	0.053	0.193
C19	0.072	0.302
C20	0.007	0.151
C21	0.009	0.152
C22	0.037	0.250
C23	0.081	0.269
C24	0	0.227
C25	0.109	0.227
C26	0.039	0.261
C27	0.031	0.224
C28	0.034	0.193
C29	0	0.239
C30	0	0.184

HRV Type[a]	Max. intra-clade distance[b]	Min. inter-clade distance[c]
A44[d]	**0.007**	**0.072**
A45	0.081	0.252
A46	0.006	0.203
A47	0.122	0.140
A49	0.100	0.141
A50	0.002	0.158
A51	0.099	0.190
A53	0.103	0.222
A54	**0.087**	**0.107**
A55	0.056	0.207
A56	0	0.201
A57	**0**	**0.129**
A58	0.086	0.157
A59	0.089	0.136
A60	0.098	0.197
A61	0.102	0.180
A62	**0.007**	**0.093**
A63	0	0.174
A64	0.006	0.158
A65	0.065	0.150
A66	0.085	0.130
B26	0.002	0.186
B27	0.003	0.176
B35	0.013	0.191
B37	0.078	0.147
B42	0.098	0.167
B48	0.110	0.188
B52	0.006	0.121
B69	0.102	0.181
B70	0.116	0.124
B72	0.099	0.186
B79	0.005	0.192
B83	0.088	0.169
B84	0.101	0.223
B86	0.105	0.218
B91	0.085	0.143
B92	0.005	0.169
B93	0.002	0.176
B97	0.004	0.194
B99	0.001	0.182
B100	0	0.166
B101	0.016	0.163
C31	0	0.219
C32	0.028	0.136
C33	0.021	0.189
C34	0.027	0.151
C35	0	0.326
C36	0.111	0.256
C37	0	0.207
C38	0.034	0.174
C39	0.026	0.136
C40	0.070	0.165
C41	0.075	0.184
C42	0.086	0.241
C43	0.116	0.247
C44	0	0.197
C45	0.043	0.239
C46	0	0.256
C47	0	0.222
C48	0	0.189
C49	0.031	0.139
C50	0	0.286
C51	0.005	0.207

[a] For ease of reference, the prefix HRV- has been omitted from list of HRV type designations. Comparisons of prototype strains which show pairwise p-distance values that violate the proposed thresholds are shown in red.

[b] Maximum VP1 p-distance observed within a bootstrap supported phylogenetic clade or putative type group.

[c] Minimum VP1 p-distance observed between one particular bootstrap supported phylogenetic clade and it's nearest neighbour

[d] In cases where contemporary strains group separately from prototype strains of two types, which should potentially be combined (A29/A44), the p-distance listed involves the prototype strains only.

4.3.3 Identification of putative new HRV types on the basis of sequence divergence in VP1

Three new types HRV-A and one of HRV-B have been previously defined on the basis of sequence divergence in VP1; HRV-A101 (Rathe et al., 2010), HRV-A102

(de Vries et al., 2008), HRV-A103 and HRV-B100 (Linsuwanon et al., 2011). In addition, the current analysis has yielded three putative new HRV-A types (tentatively labelled HRV-A104 – A106) and four novel HRV-B types (HRV-B101 – B104). These are marked in red on phylogenetic trees presented (Figure 4.3).

HRV-C sequences are now divided into 51 distinct types on the basis of sequence divergence in the VP1 region (Table 4.3). HRV-C1-33 were first defined in 2010 (Simmonds et al., 2010), while HRV-C36-46 were subsequently recognised in 2012 (McIntyre et al., 2013b). In addition, HRV-C34, C35 and C47-51 have been submitted from other sources and confirmed as distinct types by the Picornavirus Study Group (www.picornaviridae.com).

Several of the putative new HRV-A and –B types included showed a lower limit of inter-clade VP1 nucleotide divergence that was close to or below the 13% and 12% thresholds respectively (Table 4.2). For example, the 6 available sequences from the novel group putatively designated as HRV-B103 showed VP1 distances from their nearest neighbour type (HRV-B6) ranging from 0.113 to 0.148, with maximum intra-type pairwise p-distances for HRV-B6 and HRV-B103 of 0.088 and 0.035 respectively. For these two types, only two pairwise comparisons fell below the 12% threshold and both involved sequence F110_9318. When this single sequence is excluded from the analysis, the minimum value was 0.126. This discrepancy might have originated from sequencing or assembly errors, an evolutionarily intermediate sequence between HRV-B6 and HRV-B103 arising during their subsequent differentiation into distinct types or a chimaera generated by inter-type recombination of closely related types.

In contrast to the borderline VP1 divergence observed between HRV-B6 and HRV-B103, new HRV-B types -B100-102 showed pairwise p-distances in VP1 of 0.214, 0.190 and 0.227 from their nearest neighbour (HRV-B35). The novel type

designation HRV-B104 represents a division of a previously single type group (HRV-B52). This division is based on the existence of two discrete phylogenetic clades with VP1 p-distance of 0.121 to 0.142 between them. In this case, the clade containing the original ATCC isolates, including strains FJ445188 and EF173424 has retained the HRV-B52 type designation, whereas the clade containing field isolate FJ445137 and JF781506 has been designated HRV-B104.

The three putative new types detected within available HRV-A sequences consisted of two groups with borderline VP1 divergence (HRV-A104/A105) and one which had a p-distance of 0.172 from its nearest neighbour, HRV-A71 (HRV-A106). While HRV-A104 can technically be considered to be a borderline new HRV-A type, the lowest inter-clade VP1 p-distance observed between it and HRV-A24 was just above the proposed threshold (0.131). The six available HRV-A105 VP1 sequences group in a bootstrap supported clade which is closely related to, but distinct from the HRV-A57 prototype sequence, FJ445141. Pairwise VP1 p-distances range from 0.129 to 0.140 between the two groups. Only two sequences within HRV-A105 had a pairwise VP1 p-distance of below 13% with HRV-A57; AY355238 and AY450495. HRV-A105 has therefore been classified as a putative new HRV-A type with intermediate VP1 divergence.

4.3.4 Analysis of HRV type groups which do not conform to the proposed VP1 divergence thresholds

With the definition of VP1 nucleotide divergence thresholds for EV, several pairs of types showing divergence below this value were combined into single types. In many cases, this reassignment into single types was consistent with demonstrated serological cross-reactivity. As examples from human EV classification, CAV13/CAV18 and CAV11/CAV15 have been reclassified as CAV13 and CAV11 respectively (Brown et al., 2003). Analysis of pairwise VP1 divergence between

previously recognised HRV types revealed several similar examples. For example, HRV-A8 and HRV-A95 have been combined and renamed as HRV-A8, due to an inter-type VP1 p-distance range of 0.015 to 0.019 and previously documented serological cross-reactivity (Ledford et al., 2004) (Figure 4.4 A1).

Among other candidates, HRV-A29 and HRV-A44 have previously demonstrated cross-reactivity (Cooney et al., 1982) and the two clades containing the prototype strains had an inter-type VP1 p-distance of 0.0724 (Figure 4.4 A2). However, contemporary strains isolated between 2003 and 2009 grouped separately from both serotype groups and cannot be classified as either by VP1 p-distance. When both contemporary and prototype strains are considered, the A29/A44 group had a maximum divergence of 0.159 (involving the most divergent strain FJ042596). All instances of intra-clade VP1 divergence of more than 13% involved one of the two most divergent strains, FJ042596 or FJ042606 compared with another contemporary strain. However, all contemporary strains had a VP1 p-distance with the A29 and A44 prototype strains (FJ445125 and DQ473499) that is below the threshold of 13%. The most divergent strains mentioned are represented in each case by one sequence only and therefore it is not possible to determine at this stage whether these represent variants that happen to be at the extreme end of the distribution of intra-type divergence values or genuine intermediates in the process of diverging into a new HRV type. As data accumulates from further epidemiological and evolutionary studies, this may become clear. However, due to the well supported phylogenetic grouping with other strains within this group and the p-distance comparisons with prototype strains that universally fell below the threshold of 13%, this entire group of sequences has been combined and renamed as HRV-A29.

HRV-A54 and HRV-A98 are a third type pair that are more closely related than expected, with a pairwise nucleotide VP1 p-distance of 0.107 to 0.124. Unfortunately, the cross-neutralisation properties of these two serotypes have not

been analysed in previous studies (Cooney et al., 1982; Ledford et al., 2004). However, on the basis of close phylogenetic relatedness (Figure 4.4 A3) and analysis of VP1 divergence, we would suggest that these may be combined.

These three combined pairs of classical types may have been erroneous classifications in original studies (marked in yellow on Figure 4.3). In the new proposed HRV classification system, they have been assigned to a type number of the earliest submitted full genome. This therefore removes designations HRV-A44, A54 and A95. In order to avoid confusion arising from published literature referring to the now abolished type designations, we would propose retaining these gaps in the numbering scheme.

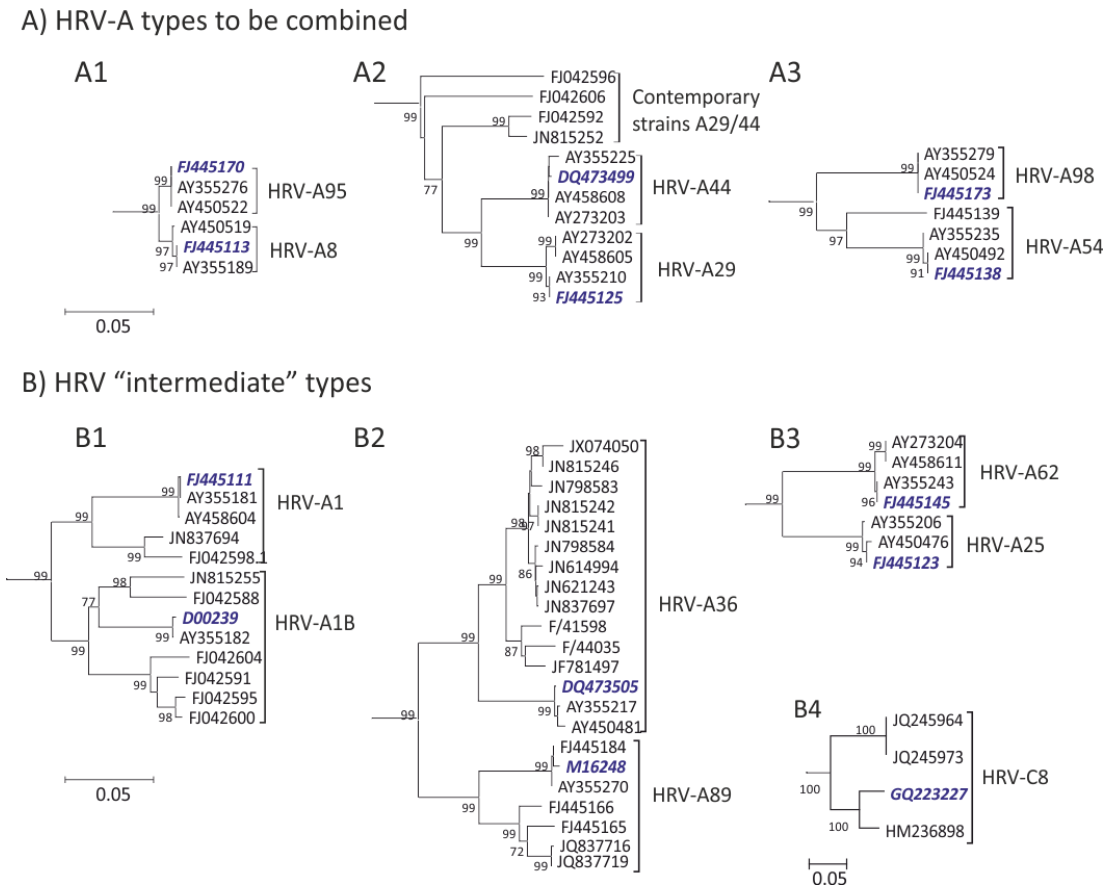


FIGURE 4.4: Neighbour joining phylogenetic trees showing classical HRV type pairs which do not conform to the proposed VP1 divergence thresholds. *For ease of reference, the branch to tree root has been collapsed and the previously recognised prototype strain of each HRV type is highlighted in blue. Three HRV-A type pairs which have been combined are shown in A1, A2 and A3. Four HRV type pairs which display intermediate divergence values are shown in B1, B2, B3 and B4. Branches are scaled by genetic distance.*

In addition to those type pairs which have been combined into single type assignments, a number of other HRV types display intermediate divergence values that violate the proposed thresholds. HRV-A1/A1B are serologically cross-reactive (Cooney et al., 1982) but fall into two discrete phylogenetic groups and have pairwise p-distances that fall above the VP1 threshold (Figure 4.4 B1). While the overall inter-clade distance ranges from 0.119 to 0.149, analysis of the prototype strains alone (A1: FJ445111 and A1B: D00239) showed a distance of 0.122. While we suggest that these two clades be considered as HRV-A1, this type designation should be reviewed as more sequence data becomes available.

As a contrasting problem, HRV-A25/A62 showed inter-clade VP1 pairwise distances ranging from 0.093 to 0.103 (Figure 4.4 B3), despite no evidence of cross-neutralisation between these two serotypes (Cooney et al., 1982). Therefore, this pair have retained their current type designation and been highlighted as an intermediate type pair. Future studies should include re-assessment of the serological relationship between these two types and their designation revised accordingly.

HRV-A36 and HRV-A89 strains fell into two distinct bootstrap supported phylogenetic clades and the prototype strains (DQ473508 and M162488) showed a VP1 p-distance of 0.129, on the borderline of being classified as the same type. However, although serum specific to HRV-A36 strains can neutralize HRV-A89, -A50 and –A58, there was no reciprocal neutralisation (Cooney et al., 1982). Analysis of all nine known strains of HRV-A36 and five known strains of HRV-A89 yielded an inter-clade VP1 divergence ranging from 0.117 to 0.172. We would

suggest considering these two type designations as distinct entities with an intermediate nucleotide divergence in VP1.

Only one HRV-C type showed p-distance values that violated proposed thresholds. HRV-C8 strains were divided into two clades, one previously classified as HRV-Cpat28 and both containing contemporary sequences (Figure 4.4 B4). The VP1 p-distance between the two clades ranged from 0.129 to 0.132. In consultation with the chairman of the Picornavirus Study Group (personal communication, N. Knowles, Picornavirus Study Group), it was decided that these two clades should remain classified as HRV-C8 on the basis of borderline VP1 divergence.

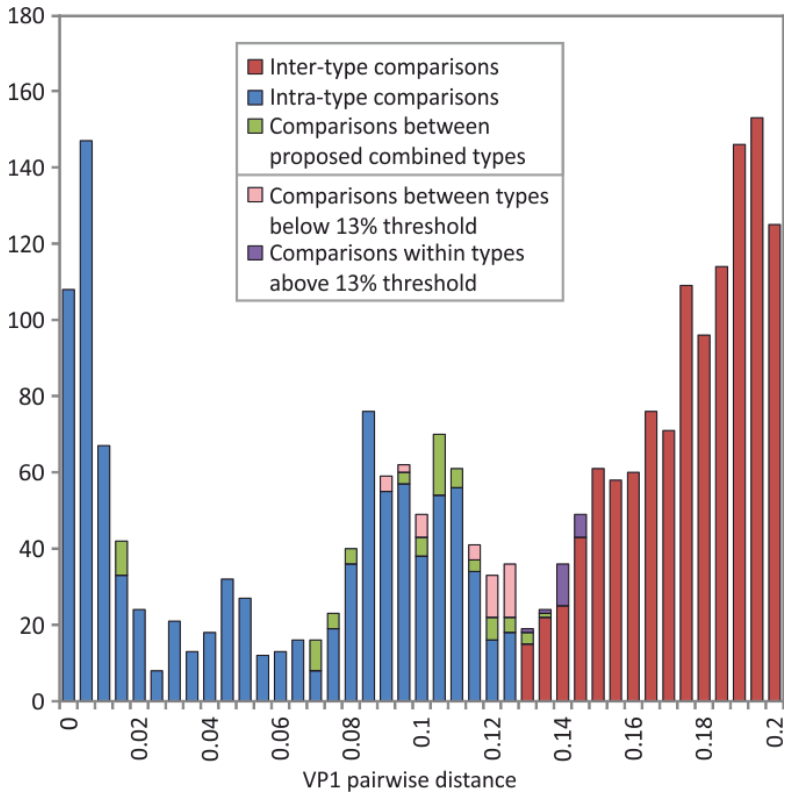


FIGURE 4.5: Distribution of HRV-A VP1 pairwise p-distances immediately surrounding the threshold of 13%. *VP1 p-distances between or within types which violate the proposed thresholds are marked in different colours. Types which should be combined are shown in green, while those which remain discrete HRV types have inter-type distances falling below the threshold are pink. Intra-type comparisons falling above the divergence threshold are shown in purple.*

Despite the large number of sequences available for all three species, HRV-A alone showed considerable evidence for VP1 pairwise p-distance comparisons that violate the proposed divergence thresholds. Of these, the majority represent comparisons between types which should be combined (labelled in green on Figure 4.5). Only 6 discrete HRV-A types showed inter-clade distances which are below the threshold (labelled in pink on Figure 4.5) and intra-clade distances which are above the threshold (labelled in purple on Figure 4.5). We would suggest the combined use of phylogenetic analysis and pairwise VP1 p-distance analysis in the classification of all HRV sequences, especially those suspected to be members of intermediate type groups.

4.3.5 Assignment of type designations based on VP1 divergence for all three species of HRV

The conclusion of the analysis of all available HRV VP1 data divided sequences into 77 HRV-A types, 29 HRV-B and 13 HRV-C on the basis of full genome sequences (Table 4.3). In addition, a further 38 HRV-C types were defined with only VP1 sequence data available (Table 4.4). The prototype sequence designated was the earliest submitted full genome or VP1 sequence (labelled in red in Table 4.3 and Table 4.4). In cases where an HRV-C type originally defined on the basis of VP1 sequence now has a full genome sequence available, the type has retained the first submitted VP1 sequence as the prototype strain (Simmonds et al., 2010). In addition, certain inconsistencies in naming exist where novel strains have been submitted by individual researchers at separate times. These have been classified based on date of submission to the Picornavirus Study Group, rather than the earliest isolated strain. These inconsistencies are unfortunately unavoidable and should remain in place for clarity.

TABLE 4.3: Confirmed HRV-A, -B and –C types with full genome sequences available

	Full Genome Sequences					
HRV Type	**Genbank Accession Number[a]**	**Strain Identifier[b]**	**Date of submission[b]**	**Full genome variants**	**VP1 variants[c]**	**Prototype strain reference**
HRV-A1	D00239 FJ445111 JN815255 JN837694	HRV 1B	1/2/00	4	14	(Hughes et al., 1988)
HRV-A2	X02316	HRV 2	14/11/06	1	4	(Skern and Sommergruber , 1985)
HRV-A7	DQ473503 FJ445176	TC-65007	27/6/07	2	4	(Kistler et al., 2007b)
HRV-A8	FJ445113[d] FJ445170	ATCC VR-1118	10/4/09	2	6	(Palmenberg et al., 2009)
HRV-A9	FJ445177[e] FJ445114 FJ445115	ATCC VR-489	10/4/09	3	10	(Palmenberg et al., 2009)
HRV-A10	DQ473498 FJ445178 JN798575 JN798582 JN541269 JN815247	TC-71602	27/6/07	6	7	(Kistler et al., 2007b)
HRV-A11	EF173414	HRV 11	15/7/07	1	3	(Tapparel et al., 2007)
HRV-A12	EF173415 HQ123441 JF781511	HRV 12	15/7/07	3	7	(Tapparel et al., 2007)
HRV-A13	FJ445116 FJ445117	ATCC VR-1123	10/4/09	2	4	(Palmenberg et al., 2009)
HRV-A15	DQ473493 JN541268	TC-66919	27/6/07	2	4	(Kistler et al., 2007b)
HRV-A16	L24917 JN562722 JN614992 JN798564 JN798574 JN815253 JN990704 JX074057	HRV 16	14/7/95	8	10	(Lee et al., 1995)
HRV-A18	FJ445118 JF781496 JF781508	ATCC VR-1128	10/4/09	3	6	(Palmenberg et al., 2009)
HRV-A19	FJ445119 JQ747746 JQ747750	ATCC VR-1129	10/4/09	3	6	(Palmenberg et al., 2009)

HRV Type	**Genbank Accession Number**	**Strain Identifier**	**Date of submission**	**Full genome variants**	**VP1 variants**	**Prototype strain reference**
HRV-A20	FJ445120 JN541270 JN614993 JN798571 JQ994494	ATCC VR-1130	10/4/09	5	8	(Palmenberg et al., 2009)
HRV-A21	FJ445121 JN837693 JQ747747	ATCC VR-1131	10/4/09	3	7	(Palmenberg et al., 2009)
HRV-A22	FJ445122	ATCC VR-1132	10/4/09	1	4	(Palmenberg et al., 2009)
HRV-A23	DQ473497 JN621244 JN815254 JN837694	TC-65847	27/6/07	4	6	(Kistler et al., 2007b)
HRV-A24	EF173416 FJ445190 JN798563	HRV 24	15/6/07	3	8	(Tapparel et al., 2007)
HRV-A25	FJ445123	ATCC VR-1135	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A28	DQ473508 JN798577 JN798580 JQ747751	TC-65852	27/6/07	4	11	(Kistler et al., 2007b)
HRV-A29	DQ473499[d] FJ445125	TC-72548	27/6/07	2	12	(Kistler et al., 2007b)
HRV-A30	DQ473512 FJ445179 JN798557	TC-67821	27/6/07	2	10	(Kistler et al., 2007b)
HRV-A31	FJ445126	ATCC VR-506	10/4/09	1	7	(Palmenberg et al., 2009)
HRV-A32	FJ445127	ATCC VR-1142	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A33	FJ445128 JN815250 JN990707	ATCC VR-330	10/4/09	3	5	(Palmenberg et al., 2009)
HRV-A34	DQ473501 FJ445189 JF781510 JF781512 JN562720	TC-65856	27/6/07	4	6	(Kistler et al., 2007b)

HRV Type	**Genbank Accession Number**	**Strain Identifier**	**Date of submission**	**Full genome variants**	**VP1 variants**	**Prototype strain reference**
HRV-A36	DQ473505 JF781497 JN614994 JN621243 JN798583 JN798584 JN815241 JN815242 JN815246 JN837697 JX074050	TC-74313	27/6/07	11	15	(Kistler et al., 2007b)
HRV-A38	DQ473495 FJ445180 JQ994496	TC-72617	27/6/07	3	5	(Kistler et al., 2007b)
HRV-A39	AY751783	HRV 39	12/4/05	1	3	(Harris and Racaniello, 2005)
HRV-A40	FJ445129 JN798579 JQ245967 JX074051	ATCC VR-341	10/4/09	4	6	(Palmenberg et al., 2009)
HRV-A41	DQ473491	TC-66066	27/6/07	1	3	(Kistler et al., 2007b)
HRV-A43	FJ445131 JN815237	ATCC VR-1153	10/4/09	2	5	(Palmenberg et al., 2009)
HRV-A45	FJ445132	ATCC VR-1155	10/4/09	1	5	(Palmenberg et al., 2009)
HRV-A46	DQ473506	TC-75908	27/6/07	1	3	(Kistler et al., 2007b)
HRV-A47	FJ445133 GQ223229 JN837692	ATCC VR-1157	10/4/09	3	10	(Palmenberg et al., 2009)
HRV-A49	DQ473496 FJ445134 JN621241 JN798561 JN798589	TC-66958	27/6/07	5	13	(Kistler et al., 2007b)
HRV-A50	FJ445135	ATCC VR-517	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A51	FJ445136 JN562725	ATCC VR-1161	10/4/09	2	4	(Palmenberg et al., 2009)
HRV-A53	DQ473507 JN798587	TC-67618	27/6/07	2	4	(Kistler et al., 2007b)
HRV-A54	FJ445138 FJ445139 FJ445173	ATCC VR-1164	10/4/09	3	7	(Palmenberg et al., 2009)

HRV Type	Genbank Accession Number	Strain Identifier	Date of submission	Full genome variants	VP1 variants	Prototype strain reference
HRV-A55	DQ473511 JQ837718	TC-64522	27/6/07	2	4	(Kistler et al., 2007b)
HRV-A56	FJ445140 EU840727	ATCC VR-1166	10/4/09	2	3	(Palmenberg et al., 2009)
HRV-A57	FJ445141	Fs ship#1	10/4/09	1	1	(Palmenberg et al., 2009)
HRV-A58	FJ445142 JX025558	ATCC VR-1168	10/4/09	2	5	(Palmenberg et al., 2009)
HRV-A59	DQ473500 JN541266	TC-70403	27/6/07	2	4	(Kistler et al., 2007b)
HRV-A60	FJ445143 JN798590	ATCC VR-1473	10/4/09	2	4	(Palmenberg et al., 2009)
HRV-A61	FJ445144 JN798560	ATCC VR-1171	10/4/09	2	6	(Palmenberg et al., 2009)
HRV-A62	FJ445145	ATCC VR-1172	10/4/09	1	4	(Palmenberg et al., 2009)
HRV-A63	FJ445146	ATCC VR-1173	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A64	EF173417 FJ445181	HRV 64	15/7/07	2	4	(Tapparel et al., 2007)
HRV-A65	FJ445147 JF781504 JQ245966	ATCC VR-1175	10/4/09	3	5	(Palmenberg et al., 2009)
HRV-A66	FJ445148 JN112340 JN621246 JQ837715	ATCC VR-1176	10/4/09	4	6	(Palmenberg et al., 2009)
HRV-A67	FJ445149 JN621245	ATCC VR-1177	10/4/09	1	6	(Palmenberg et al., 2009)
HRV-A68	FJ445150 JN798578	ATCC VR-1178	10/4/09	2	4	(Palmenberg et al., 2009)
HRV-A71	FJ445152	ATCC VR-1181	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A73	DQ473492	TC-73172	27/6/07	1	3	(Kistler et al., 2007b)
HRV-A74	DQ473494	TC-70882	27/6/07	1	3	(Kistler et al., 2007b)
HRV-A75	DQ473510 JF781503 JN837690	TC-70308	27/6/07	3	5	(Kistler et al., 2007b)
HRV-A76	DQ473502 FJ445182 JN815238 JX074049 JX074055	TC-70326	27/6/07	5	8	(Kistler et al., 2007b)

HRV Type	**Genbank Accession Number**	**Strain Identifier**	**Date of submission**	**Full genome variants**	**VP1 variants**	**Prototype strain reference**
HRV-A77	FJ445154	ATCC VR-1187	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A78	EF173418 FJ445183	HRV 78	15/7/07	2	5	(Tapparel et al., 2007)
HRV-A80	FJ445156 JN798586 JN798576 JN990705	ATCC VR-1190	10/4/09	4	6	(Palmenberg et al., 2009)
HRV-A81	FJ445157 FJ445158 FJ445159 HQ123442	ATCC VR-1191	10/4/09	4	6	(Palmenberg et al., 2009)
HRV-A82	DQ473509 FJ445160 JN798556 JN798585 JQ837722	HRV A complete genome	27/6/07	5	7	(Kistler et al., 2007b)
HRV-A88	DQ473504	TC-70782	27/6/07	1	3	(Kistler et al., 2007b)
HRV-A89	M16248 FJ445165 FJ445166 FJ445184 JQ837716 JQ837719	HRV 89	7/2/03	6	7	(Duechler et al., 1987)
HRV-A90	FJ445167	ATCC VR-1291	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A94	EF173419 FJ445185	HRV 94	15/7/07	2	4	(Tapparel et al., 2007)
HRV-A96	FJ445171	ATCC VR-1296	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A100	FJ445175	ATCC VR-1300	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-A101	GQ415051 GQ415052 JQ245965	HRV-A101	13/1/10	3	3	(Rathe et al., 2010)
HRV-A102	EF155421	AMS323	31/7/10	1	1	(de Vries et al., 2008)
HRV-A103	JF965515 JQ747749 JQ994499	WA327E /09	2/6/11	3	3	Unpub.
HRV-A104	JN562727 JX074047 JX193797	P1025 sR2625 2009	1/8/12	3	3	Unpub.

HRV Type	**Genbank Accession Number**	**Strain Identifier**	**Date of submission**	**Full genome variants**	**VP1 variants**	**Prototype strain reference**
HRV-A105	JN614995 JN990699	P1064 sR985 2009	1/8/12	2	2	Unpub.
HRV-A106	JQ245971 JX025555[f]	P1042 sR141 2008	16/5/12	2	2	Unpub.
HRV-B3	DQ473485 EF173422	TC-64701	27/6/07	2	5	(Kistler et al., 2007b)
HRV-B4	DQ473490 JN798573	HRV4	27/6/07	2	4	(Kistler et al., 2007b)
HRV-B5	FJ445112	ATCC VR-485	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B6	DQ473486 JN562723 JQ747745 JQ747748 JNX193795	TC-65006	27/6/07	5	8	(Kistler et al., 2007b)
HRV-B14	L05355 K02121 X01087	HRV 14	11/6/93	3	4	(Lee et al., 1993)
HRV-B17	EF173420	HRV 17	15/7/07	1	3	(Tapparel et al., 2007)
HRV-B26	FJ445124	ATCC VR-1136	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B27	EF173421 FJ445186	HRV 27	15/7/07	2	3	(Tapparel et al., 2007)
HRV-B35	DQ473487 FJ445187	TC-73280	27/6/07	2	4	(Kistler et al., 2007b)
HRV-B37	EF173423	HRV 37	15/7/07	1	4	(Tapparel et al., 2007)
HRV-B42	FJ445130 JN562724 JN781498 JN781507	ATCC VR-338	10/4/09	4	6	(Palmenberg et al., 2009)
HRV-B48	DQ473488 JN990698	TC-70326	27/6/07	2	4	(Kistler et al., 2007b)
HRV-B52	EF173424 FJ445188	HRV 52	15/7/07	2	4	(Tapparel et al., 2007)
HRV-B69	FJ445151 HQ123445 JN562721 JQ245970	ATCC VR-1179	10/4/09	4	10	(Palmenberg et al., 2009)
HRV-B70	DQ473489 JQ245974	TC-72589	27/6/07	2	6	(Kistler et al., 2007b)

HRV Type	Genbank Accession Number	Strain Identifier	Date of submission	Full genome variants	VP1 variants	Prototype strain reference
HRV-B72	FJ445153 GU968948 JN562726 JN614997 JN798562 JQ245969	ATCC VR-1182	10/4/09	6	8	(Palmenberg et al., 2009)
HRV-B79	FJ445155	ATCC VR-1189	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B83	FJ445161 JN990701	ATCC VR-1193	10/4/09	2	4	(Palmenberg et al., 2009)
HRV-B84	FJ445162 JF781499 JF781502 JN541271 JN614991 JQ837723 JX074048	ATCC VR-1194	10/4/09	7	9	(Palmenberg et al., 2009)
HRV-B86	FJ445164	ATCC VR-1196	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B91	FJ445168	ATCC VR-1292	10/4/09	1	4	(Palmenberg et al., 2009)
HRV-B92	FJ445169	ATCC VR-1293	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B93	EF173425	HRV 93	15/7/07	1	3	(Tapparel et al., 2007)
HRV-B97	FJ445172	ATCC VR-1297	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B99	FJ445174	ATCC VR-1299	10/4/09	1	3	(Palmenberg et al., 2009)
HRV-B100	HQ123444	CU211	23/1/11	1	1	(Linsuwanon et al., 2011)
HRV-B101	JF781500 JF781501 JX074052	P1156 sR810 2007	1/8/12	3	4	Unpub.
HRV-B102	JX074052	P1044 sR122 2007	8/6/12	1	1	Unpub.
HRV-B103	JN614996 JN798572 JQ245972 JQ837717 JQ994497	P1160 sR1153 2009	1/8/12	5	6	Unpub.
HRV-B104	FJ445137 JF781506	F10	10/4/09	2	4	(Palmenberg et al., 2009)
HRV-C1	EF077279 HQ123443	NAT001	20/10/06	2	5	(Kistler et al., 2007a)

HRV Type	Genbank Accession Number	Strain Identifier	Date of submission	Full genome variants	VP1 variants	Prototype strain reference
HRV-C2	EF077280 JN815248 JN837695 JN990703 JQ245968 JX025557	NAT043	20/10/06	6	6	(Kistler et al., 2007a)
HRV-C3	EF186077 JN798567 JN990700	QPM	14/12/06	3	22	(McErlean et al., 2007)
HRV-C4	EF582385 JF781509	024	27/4/07	2	2	(Lau et al., 2007)
HRV-C5	EF582386	025	27/4/07	1	3	(Lau et al., 2007)
HRV-C6	EF582387 JF317016 JN990702	026	27/4/07	3	14	(Lau et al., 2007)
HRV-C7	DQ875932 JN798559 JN798570 JN837689 JX025556	NY-074	14/7/08	5	7	(Lamson et al., 2006)
HRV-C8	GQ223227 JQ245964 JQ245973	N4	29/5/09	3	4	(Huang et al., 2009)
HRV-C9	GQ223228	N10	29/5/09	1	8	(Huang et al., 2009)
HRV-C10	GQ323774	QCE	29/6/09	1	6	(Arden et al., 2010)
HRV-C11	EU840952	CL-170085	21/5/10	1	1	(Tapparel et al., 2009a)
HRV-C35	JF436925	Subtype 35	13/6/12	1	1	(Lysholm et al., 2012)
HRV-C51	JF317015 JX291115	LZ508		2	2	Unpub.

[a] Designated prototype strain is shown in red.
[b] Strain identifier and date of submission both refer to prototype strain.
[c] Number of VP1sequences available including full genomes.
[d] Where some classical HRV type groups have been combined, the new prototype strain given is the earliest isolated full genome of either classical type.
[e] In cases where more than one full genome has been sequenced and submitted concurrently, the full genome derived from amplification of an original ATCC strain has been designated as the prototype strain.
[f] JX025555 has been designated as the prototype strain of HRV-A104, as the other available polyprotein sequence JQ245971 has a large gap within the VP1 region.

TABLE 4.4: Confirmed HRV–C types defined by VP1 sequence

	Full Genome Sequences	**VP1 sequences**				
HRV Type	**Genbank Accession Number**	**Genbank Accession Number**	**Strain Identifier**	**Date of submission**	**VP1 variants**	**Prototype strain reference**
HRV-C12	JF317017	HM236958	Resp_3922/07	14/5/10	7	(McIntyre et al., 2010)
HRV-C13	-	HM236908	Resp_2951/06	14/5/10	2	(McIntyre et al., 2010)
HRV-C14	-	HM236911	Resp_3090/06	14/5/10	1	(McIntyre et al., 2010)
HRV-C15	GU219984 JF317014 JN837688	HM236963	Resp_4644/07	14/5/10	4	(McIntyre et al., 2010)
HRV-C16	-	HM236944	Resp_5910/07	14/5/10	3	(McIntyre et al., 2010)
HRV-C17	JN815240 JN815244	HM236936	Resp_5145/07	14/5/10	3	(McIntyre et al., 2010)
HRV-C18	-	HM236918	Resp_3631/07	14/5/10	6	(McIntyre et al., 2010)
HRV-C19	-	EU840728	CL-Fnp5	20/6/08	3	(Tapparel et al., 2009a)
HRV-C20	-	HM236923	Resp_3995/07	14/5/10	2	(McIntyre et al., 2010)
HRV-C21	-	HM236903	Resp_5071/07	14/5/10	2	(McIntyre et al., 2010)
HRV-C22	JN621242	HM236905	Resp_2748/06	14/5/10	3	(McIntyre et al., 2010)
HRV-C23	-	HM236901	Resp_3053/06	14/5/10	2	(McIntyre et al., 2010)
HRV-C24	-	HM236939	Resp_7147/07	14/5/10	1	(McIntyre et al., 2010)
HRV-C25	HQ123440 JF317013 JN837685	HM236952	Resp_2832/06	14/5/10	4	(McIntyre et al., 2010)
HRV-C26	JX193796	HM236904	Resp_2514/06	14/5/10	3	(McIntyre et al., 2010)
HRV-C27	-	HM236906	Resp_2784/06	14/5/10	4	(McIntyre et al., 2010)
HRV-C28	JN798569	HM236954	Resp_3105/06	14/5/10	2	(McIntyre et al., 2010)
HRV-C29	-	HM236949	Resp_5345/07	14/5/10	1	(McIntyre et al., 2010)
HRV-C30	-	HM236968	Resp_3898/07	14/5/10	1	(McIntyre et al., 2010)

HRV Type	**Genbank Accession Number**	**Genbank Accession Number**	**Strain Identifier**	**Date of submission**	**VP1 variants**	**Prototype strain reference**
HRV-C31	-	HM236964	Resp_4923/07	14/5/10	1	(McIntyre et al., 2010)
HRV-C32	JN798581 JQ994498	HM236897	Resp_6131/07	14/5/10	12	(McIntyre et al., 2010)
HRV-C33	-	HM236934	Resp_4917/07	14/5/10	2	(McIntyre et al., 2010)
HRV-C34	JF436926	JF519798	PNG7111-3625	22/6/12	2	(Chidlow et al., 2012)
HRV-C36	JN541267	JF416311	Resp_2480/07	22/8/11	5	(McIntyre et al., 2013b)
HRV-C37	-	JF416321	Resp_6135/08	22/8/11	1	(McIntyre et al., 2013b)
HRV-C38	JN837691	JF416322	Resp_6142/08	22/8/11	2	(McIntyre et al., 2013b)
HRV-C39	-	JF416306	Resp_10221/08	22/8/11	2	(McIntyre et al., 2013b)
HRV-C40	JF781505 JN815251	JF416312	Resp_2800/06	22/8/11	6	(McIntyre et al., 2013b)
HRV-C41	JN798565	JF416323	Resp_9449/08	22/8/11	3	(McIntyre et al., 2013b)
HRV-C42	JQ994500	JF416320	Resp_5477/08	22/8/11	2	(McIntyre et al., 2013b)
HRV-C43	JN837687 JX074056	JF416307	Resp_13229/08	22/8/11	4	(McIntyre et al., 2013b)
HRV-C44	-	JF416310	Resp_15588/09	22/8/11	1	(McIntyre et al., 2013b)
HRV-C45	JN837686	JF416308	Resp_13958/08	22/8/11	3	(McIntyre et al., 2013b)
HRV-C46	-	JF416318	Resp_5153/07	22/8/11	2	(McIntyre et al., 2013b)
HRV-C47	-	JF519760	PNG7254-3947	22/6/12	1	(Chidlow et al., 2012)
HRV-C48	-	JF519762	PNG7293-3193	22/6/12	1	(Chidlow et al., 2012)
HRV-C49	JF907574	JF946738	WA11040M	22/10/11	3	(Wiertsema et al., 2011)
HRV-C50	-	JQ739202	SG1	24/4/12	1	Unpub.

4.3.6 Analysis of recombination within the capsid coding region of HRV and phylogenetic analysis of the VP4/VP2 region

Within EV isolates, recombination has been documented within the capsid coding region (Bouslama et al., 2007) and therefore the VP1 region alone is deemed suitable for typing. However, recombination within the capsid coding region of HRV is thought to be relatively infrequent (Palmenberg et al., 2009; Kistler et al., 2007b; Lewis-Rogers et al., 2009).

The well conserved and readily amplified VP4/VP2 region of HRV is commonly used in studies of its epidemiology and clinical associations. Recent and rapid accumulation of sequence data has created over 3900 HRV sequences in the region available on GenBank compared to 384 sequences of the VP1 region, a region that is problematic to amplify without species or type-specific primers. Therefore, in view of this imbalance, it is important to verify that known HRV types can be correctly identified through analysis of the VP4/VP2 region if sequence data from this region is to be employed in epidemiological and clinical studies.

The majority of HRV strains grouped congruently between VP1 and VP4/VP2 (data not shown), indicating that phylogenetic grouping in VP4/VP2 was predictive of the type group assigned using VP1 sequences. Only one sequence showed evidence of phylogenetic incongruence within the capsid region. JX291115 is a member of HRV-C51 type group in the VP1 region, with a pairwise nucleotide p-distance of 0.005. However, in VP4/VP2, this sequence had a VP1 p-distance of 0.1459 from HRV-C51 and groups as a poorly supported outgroup to HRV-C15. Analysis with the RDP software package confirms this sequence as a putative recombinant and similar to the pattern observed from phylogenetic analysis, there was no known minor recombination parent. Analysis of HRV-A and -B carried out with RDP and

GARD highlighted no evidence of recombination between regions. This suggests that the VP4/VP2 region may additionally be used for classification of HRV types.

We previously proposed a divergence threshold of 10% for the purpose of classifying HRV-C sequences by VP4/VP2 sequence (Simmonds et al., 2010). This was calculated using all HRV sequences that were over 90% complete across a fragment of VP4/VP2 from position 615 – 1043 (numbered by EF582385). In the revised analysis, the use of a shorter fragment was considered (615 – 1004 numbered by EF582385), as this would allow inclusion of a large number of additional VP4/VP2 sequences. Inspection of phylogenetic trees constructed for both fragments showed no change in the uniformly high bootstrap support for individual type groups and no change in the clade membership of individual sequences. Therefore, the smaller fragment was used in all subsequent phylogenetic and p-distance analysis (Figure 4.6).

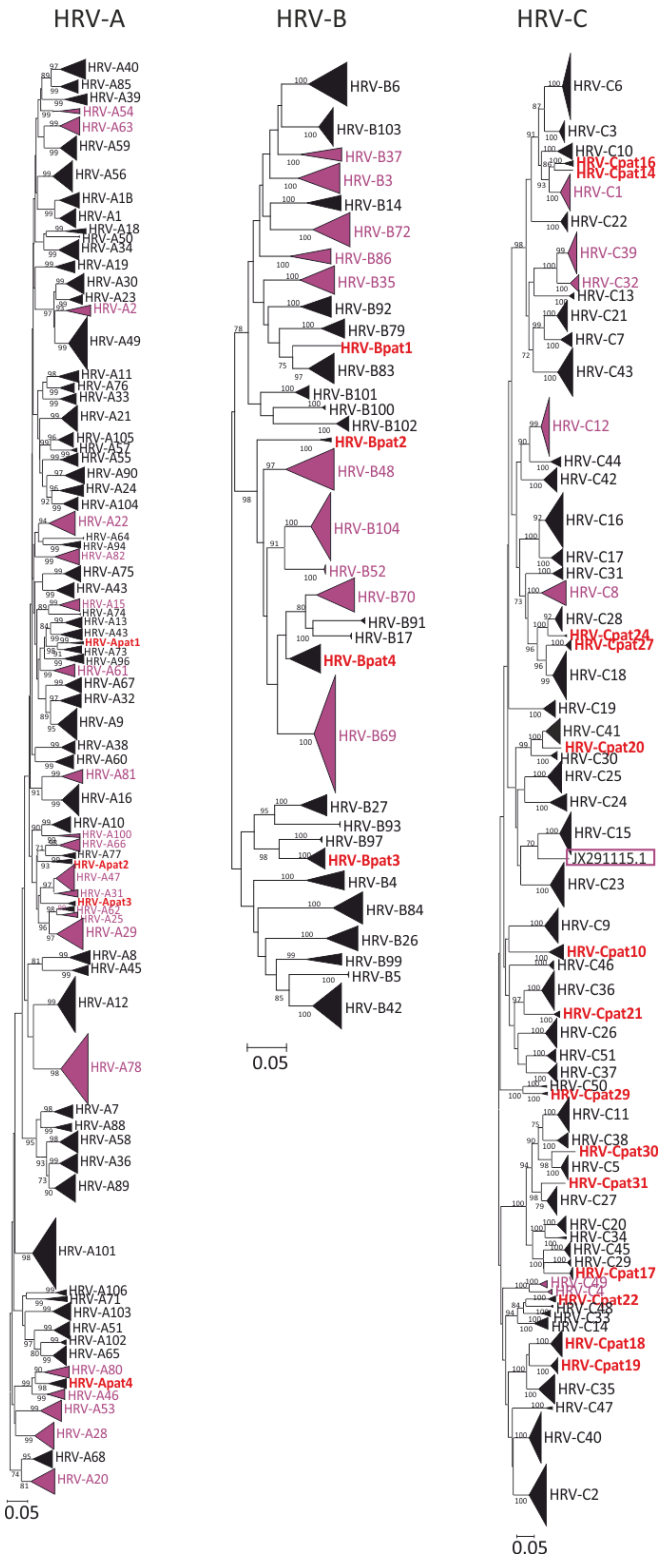

HRV-A
HRV-A40
HRV-A85
HRV-A39
HRV-A54
HRV-A63
HRV-A59
HRV-A56
HRV-A1B
HRV-A1
HRV-A18
HRV-A50
HRV-A34
HRV-A19
HRV-A30
HRV-A23
HRV-A2
HRV-A49
HRV-A11
HRV-A76
HRV-A33
HRV-A21
HRV-A105
HRV-A57
HRV-A55
HRV-A90
HRV-A24
HRV-A104
HRV-A22
HRV-A64
HRV-A94
HRV-A82
HRV-A75
HRV-A43
HRV-A15
HRV-A74
HRV-A13
HRV-A43
HRV-Apat1
HRV-A73
HRV-A96
HRV-A61
HRV-A67
HRV-A32
HRV-A9
HRV-A38
HRV-A60
HRV-A81
HRV-A16
HRV-A10
HRV-A100
HRV-A66
HRV-A77
HRV-Apat2
HRV-A47
HRV-A31
HRV-Apat3
HRV-A62
HRV-A25
HRV-A29
HRV-A8
HRV-A45
HRV-A12
HRV-A78
HRV-A7
HRV-A88
HRV-A58
HRV-A36
HRV-A89
HRV-A101
HRV-A106
HRV-A71
HRV-A103
HRV-A51
HRV-A102
HRV-A65
HRV-A80
HRV-Apat4
HRV-A46
HRV-A53
HRV-A28
HRV-A68
HRV-A20
0.05
HRV-B
HRV-B6
HRV-B103
HRV-B37
HRV-B3
HRV-B14
HRV-B72
HRV-B86
HRV-B35
HRV-B92
HRV-B79
HRV-Bpat1
HRV-B83
HRV-B101
HRV-B100
HRV-B102
HRV-Bpat2
HRV-B48
HRV-B104
HRV-B52
HRV-B70
HRV-B91
HRV-B17
HRV-Bpat4
HRV-B69
HRV-B27
HRV-B93
HRV-B97
HRV-Bpat3
HRV-B4
HRV-B84
HRV-B26
HRV-B99
HRV-B5
HRV-B42
0.05
HRV-C
HRV-C6
HRV-C3
HRV-C10
HRV-Cpat16
HRV-Cpat14
HRV-C1
HRV-C22
HRV-C39
HRV-C32
HRV-C13
HRV-C21
HRV-C7
HRV-C43
HRV-C12
HRV-C44
HRV-C42
HRV-C16
HRV-C17
HRV-C31
HRV-C8
HRV-C28
HRV-Cpat24
HRV-Cpat27
HRV-C18
HRV-C19
HRV-C41
HRV-Cpat20
HRV-C30
HRV-C25
HRV-C24
HRV-C15
JX291115.1
HRV-C23
HRV-C9
HRV-Cpat10
HRV-C46
HRV-C36
HRV-Cpat21
HRV-C26
HRV-C51
HRV-C37
HRV-C50
HRV-Cpat29
HRV-C11
HRV-C38
HRV-Cpat30
HRV-C5
HRV-Cpat31
HRV-C27
HRV-C20
HRV-C34
HRV-C45
HRV-C29
HRV-Cpat17
HRV-Cpat22
HRV-C48
HRV-C33
HRV-C14
HRV-Cpat18
HRV-Cpat19
HRV-C35
HRV-C47
HRV-C40
HRV-C2
0.05

FIGURE 4.6: Neighbour joining phylogenetic trees showing the VP4/VP2 region of all three species of HRV. *For ease of reference, individual type groups and the branch to tree root have been collapsed. Divergent groups which have been temporarily designated as provisionally assigned types are shown in red. HRV type groups which do not conform to the proposed VP4/VP2 thresholds are shown in purple. The single recombinant sequence detected, JX291115 is marked by a purple box. Branches are scaled by genetic distance.*

4.3.7 Assigning nucleotide divergence thresholds for the VP4/VP2 region

Pairwise p-distance comparisons are divided into two distributions in the VP4/VP2 region (Figure 4.7). Detailed inspection of distributions of pairwise distances allowed a threshold of 10.5%, 9.5% and 10.5% to be assigned for HRV-A, -B and –C respectively. Due to the fact that VP4/VP2 is a considerably shorter and less divergent sequence fragment than VP1, the thresholds are less clearly defined, with a greater number of pairwise comparisons close to the assignment threshold.

The original HRV-C classification included 28 provisionally assigned types, which were assigned on the basis of VP4/VP2 sequence divergence alone. When a corresponding VP1 sequence was obtained and confirmed to display at least 13% divergence, a new confirmed type was created and the provisionally assigned type designation discarded. Of these original 28 provisionally assigned types, only 11 remain (listed in Table 4.5). In 2012, VP1 sequencing of HRV-C isolates from 10 provisionally assigned type groups led to confirmation of further species C types HRV-C36 to HRV-C46 (McIntyre et al., 2013b). To date, only one provisionally assigned type has been combined into an existing type (HRV-C8/Cpat28). This reaffirms the usefulness of the provisionally assigned type system for identifying candidate new types and guiding the use of VP1 region sequencing for confirming these.

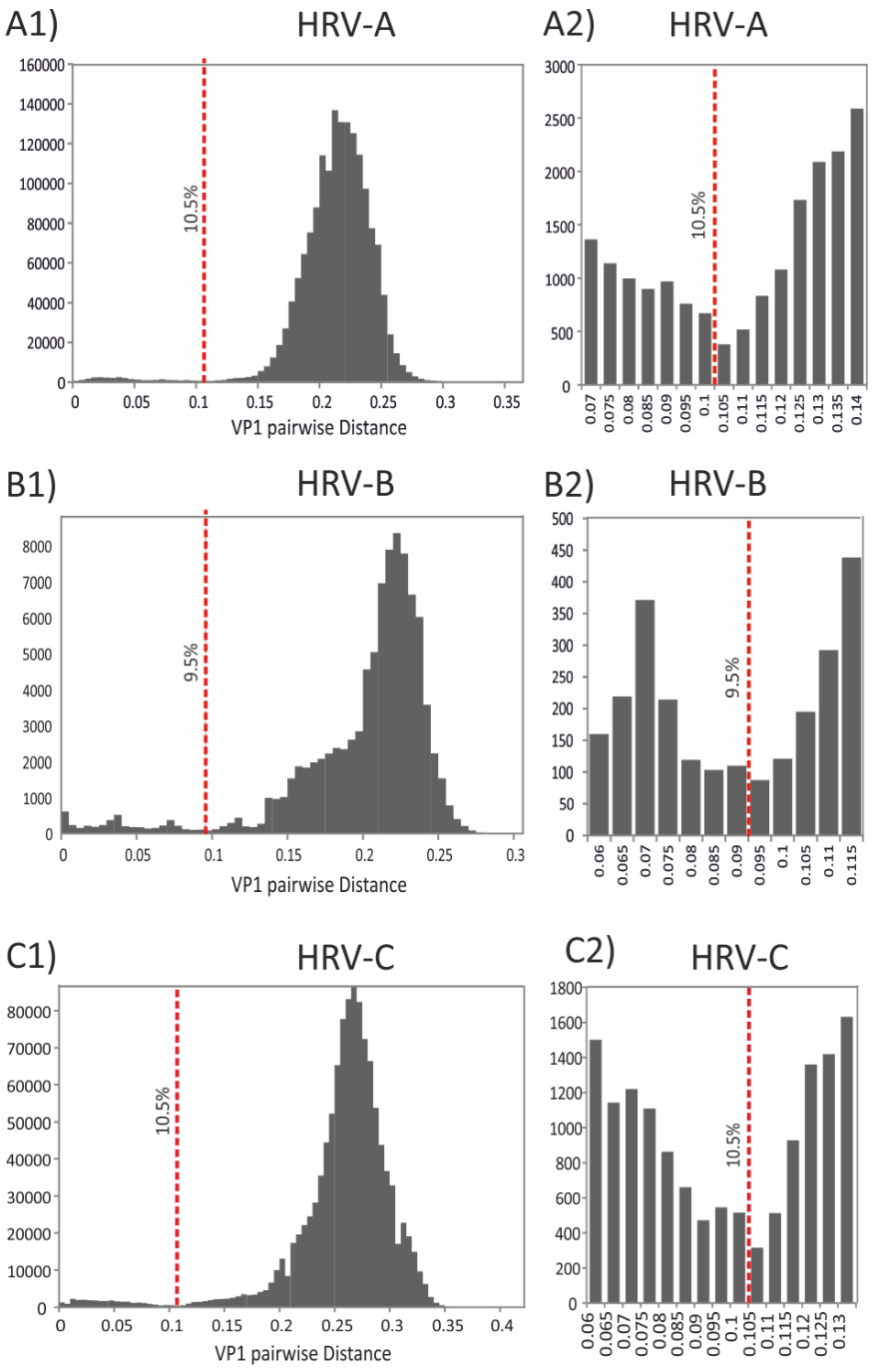


FIGURE 4.7: ***Distributions of pairwise nucleotide p-distances for the VP4/VP2 region of all three species of HRV.*** *HRV-A is shown in section A1, HRV-B in B1 and HRV-C in C1. A detailed analysis of the distribution values adjacent to the proposed thresholds for HRV-A (A2), HRV-B (B2) and HRV-C (C2) is also shown and the proposed nucleotide divergence threshold is shown by a red dotted line.*

Based on this principle, 4 previously unrecognised divergent VP4/VP2 groups were observed in HRV-A and 4 in HRV-B (labelled in red on Figure 4.6). A further three provisionally assigned HRV-C types were also detected (designated HRV-Cpat29-pat31). All showed divergence above the VP4/VP2 threshold proposed for their species (Table 4.5). Only one of these provisionally assigned types, HRV-Apat2 had a corresponding sample which existed within our archive. However, the sample was missing and so VP1 sequencing could not be attempted.

TABLE 4.5: Provisionally assigned types of HRV based on VP4/VP2 sequence divergence

HRV Species	HRV p.a.t	VP4/VP2 type strain	Submission date	VP4/ VP2 strains	Closest type[a]	p-dist[b]	Type strain ref
A	Apat1	GU568105	13/10/10	2	A73	0.106	(Xiang et al., 2010b)
A	Apat2	GQ476621	4/12/09	4	A47	0.119	(Wisdom et al., 2009b)
A	Apat3	AB549407	10/3/10	6	A43	0.144	(Kaida et al., 2011)
A	Apat4	EU590059	24/6/09	12	A46	0.112	(Savolainen-Kopra et al., 2009a)
B	Bpat1	EF077241	23/8/07	1	B83	0.119	(Kistler et al., 2007a)
B	Bpat2	HM366914	3/1/11	2	B48	0.122	(Piralla et al., 2011)
B	Bpat3	HM366910	3/1/11	11	B97	0.097	(Piralla et al., 2011)
B	Bpat4	AB548901	10/3/10	15	B91	0.107	Unpub.
C	Cpat10	EU590054	25/3/08	17	C2	0.200	(Savolainen-Kopra et al., 2009a)
C	Cpat14	EU697852	5/5/08	6	Cpat16	0.106	(Briese et al., 2008)
C	Cpat16	EU752358	26/5/08	5	Cpat14	0.106	(Miller et al., 2009a)
C	Cpat17	EU752398	26/5/08	14	C29	0.133	(Miller et al., 2009a)
C	Cpat18	EU752412	26/5/08	28	Cpat19	0.146	(Miller et al., 2009a)
C	Cpat19	FJ598096	31/12/09	17	Cpat18	0.146	Unpub.
C	Cpat20	FJ615722	9/1/09	1	C41	0.105	(Miller et al., 2009b)
C	Cpat21	FJ615737	9/1/09	7	C36	0.144	(Miller et al., 2009b)
C	Cpat22	FJ615745	9/1/09	7	C33	0.121	(Miller et al., 2009b)
C	Cpat24	FJ869923	27/3/09	2	C28	0.106	(Han et al., 2009)
C	Cpat27	GU214340	18/11/09	10	C18	0.105	(Piralla et al., 2009)
C	Cpat29	FR820909	3/4/12	3	C50	0.113	(Henquell et al., 2012)
C	Cpat30	AB550405	23/12/11	1	C5	0.110	(Fujitsuka et al., 2011)
C	Cpat31	AB628117	24/2/12	1	C27	0.115	Unpub.

[a] This represents the closest HRV type by p-distance and phylogenetic analysis.

[b] Lowest inter-clade VP4/VP2 pairwise nucleotide p-distance between provisionally assigned type and nearest neighbour

4.3.8 Analysis of confirmed HRV types which do not conform to VP4/VP2 thresholds

Analysis of inter- and intra-type pairwise nucleotide p-distances in VP4/VP2 revealed many groups which violated the proposed divergence thresholds (22, 10 and 6 types of HRV-A, -B and –C respectively). HRV-A1 and HRV-A25/62 displayed the same patterns as in the VP1 region, in which they were designated as intermediate types.

TABLE 4.6: Limits of inter- and intra- clade divergence in HRV types which do not conform to VP4/VP2 thresholds

HRV Type	Prototype strain accession number	VP4/VP2 variants	VP4/VP2 intra-clade comparisons upper limit	VP4/VP2 inter-clade comparisons lower limit
HRV-A1[a]	***D00239***	***44***	***0.124***	***0.124***
HRV-A2	***X02316***	***15***	***0.123***	***0.133***
HRV-A15	DQ473493	17	0.107	0.147
HRV-A20	***FJ445120***	***23***	***0.119***	***0.119***
HRV-A22	***FJ445122***	***13***	***0.136***	***0.148***
HRV-A25	FJ445123	7	0.075	0.064
HRV-A28	***DQ473508***	***30***	***0.114***	***0.151***
HRV-A29[b]	***DQ473499***	***42***	***0.159***	***0.151***
HRV-A31	***FJ445126***	***10***	***0.097***	***0.083***
HRV-A46	DQ473506	14	0.137	0.119
HRV-A47	***FJ445133***	***39***	***0.098***	***0.083***
HRV-A53	DQ473507	28	0.109	0.128
HRV-A54	***FJ445138***	***12***	***0.129***	***0.125***
HRV-A61	FJ445144	17	0.112	0.135
HRV-A62	FJ445145	7	0.064	0.064
HRV-A63	FJ445146	24	0.118	0.136
HRV-A66	***FJ445148***	***16***	***0.112***	***0.130***
HRV-A78	***EF173418***	***83***	***0.153***	***0.163***
HRV-A80	***FJ445156***	***15***	***0.129***	***0.129***
HRV-A81	***FJ445157***	***17***	***0.107***	***0.152***
HRV-A82	***DQ473509***	***20***	***0.135***	***0.147***
HRV-A100	FJ445175	5	0.143	0.147
HRV-B3	U60874	13	0.128	0.142
HRV-B35	DQ473487	13	0.102	0.135
HRV-B37	***EF173423***	***6***	***0.102***	***0.118***
HRV-B48	***DQ473488***	***22***	***0.129***	***0.122***
HRV-B52	***EF173424***	***32***	***0.005***	***0.093***
HRV-B69	FJ445151	61	0.131	0.099

HRV Type	Prototype strain accession number	VP4/VP2 variants	VP4/VP2 intra-clade comparisons upper limit	VP4/VP2 inter-clade comparisons lower limit
HRV-B70	***DQ473489***	***16***	***0.100***	***0.097***
HRV-B72	FJ445153	17	0.102	0.151
HRV-B86	FJ445164	7	0.107	0.138
HRV-B104	***FJ445137***	***5***	***0.092***	***0.093***
HRV-C4	***EF582385***	***6***	***0.031***	***0.100***
HRV-C8	***GQ223227***	***25***	***0.147***	***0.169***
HRV-C12	HM236958	67	0.113	0.134
HRV-C32	***HM236897***	***19***	***0.071***	***0.090***
HRV-C39	***JF416306***	***46***	***0.074***	***0.090***
HRV-C49	***JF946738***	***8***	***0.049***	***0.100***

[a] Confirmed types which are distinct in the VP1 region but violate the divergence thresholds in VP4/VP2 are marked in bold italics.

[b] Please note that new HRV type designations as presented in Table 4.3 are used in this analysis.

Four type pairs had VP4/VP2 divergence that was below the proposed threshold. HRV-A31/A47, HRV-B52/B104, HRV-C4/C49 and HRV-C32/C39 had inter-type p-distances in VP4/VP2 of 0.083, 0.093, 0.100 and 0.090 respectively. In each case, VP1 divergence between the two members of the pair confirmed them to be distinct types and each type group forms a distinct phylogenetic cluster which is readily distinguishable from its nearest neighbour. These represent confirmed type groups which are distinct in the VP1 region, but less divergent than would be expected in VP4/VP2 (marked in bold italics on Table 4.6).

Most type groups contained two sub-clusters; one containing older strains closely related to the prototype and the other containing contemporary isolates. The majority of type groups with intra-type VP4/VP2 distances which violated the thresholds had a representative VP1 sequence from the divergent contemporary cluster. Invariably, these fell within the proposed thresholds for divergence in the VP1 region. These types represent confirmed type groups which are more divergent in VP4/VP2 than would be expected (marked in bold italics on Table 4.6). Each of these types formed a distinct phylogenetic cluster in VP4/VP2.

The remaining 12 HRV types showed divergence above the threshold in VP4/VP2 and had no available VP1 sequence to confirm the type designation. These could be categorised into two groups. For the majority of these types (including HRV-A15, -A53, -A63, -A100, -B35, -B69, -B72 and –C12), all p-distance comparisons above the threshold occurred between contemporary isolates. When considered individually, every contemporary strain had a p-distance from the prototype strain of less than 10.5% or 9.5%. Therefore, these types contained examples of divergent contemporary strains that were still type group members. The remaining types could not be definitively classified. HRV-A46, -A61, -A86 and –B3 all contained at least one contemporary strain that had a pairwise nucleotide p-distance of greater than 10.5% or 9.5% with their corresponding prototype strain. However, in each case, phylogenetic analysis revealed that the divergent strain was embedded within a clade containing type group members.

The existence of such a large number of HRV types which violate the proposed thresholds highlights the fact that it is not sufficient to rely on pairwise nucleotide p-distance analysis for identification of HRV types. However, when considered in parallel with phylogenetic analysis, all strains could be readily classified into types with a degree of certainty.

4.3.9 Geographical distribution of HRV types

An analysis of the broad geographical distribution of HRV types was undertaken by obtaining information on geographical location of isolation of all VP4/VP2 sequences, where available. VP4/VP2 sequences were selected in preference over VP1 sequences, as this region is the most commonly used in prevalence studies of HRV. Although most sequences had associated geographical data available, 492 HRV isolates could not be associated with a particular location.

Although there is clearly a finite limit to the number of currently circulating HRV types, the continuing discovery of novel HRV types suggests that the limit has not yet been reached. HRV screening has been undertaken over a wide geographical base including Europe, USA, Australia and Asia. For the purposes of this analysis, sequences were divided into those isolated in 4 large regions; USA, Asia, Europe and other (including Australia, Jordan and Africa). As a substantial number of the sequences included were derived from our own screening, sequences from the UK and those from other European studies were separated into two distinct categories. The distribution of all identified strains between the geographical regions was highly similar across all three species, which indicates a lack of bias in the transmission and circulation of any HRV species within a particular geographical region (Figure 4.8).

Several HRV types were excluded from the analysis, on the basis of limited availability of geographical information (HRV-A50, -A57, -A64, -B5, -B17- B91 and –C50). All of these types consisted of one or two VP4/VP2 sequences. In addition, 5 HRV-C and 2 HRV-B types were represented by one available strain only. Without further large-scale epidemiological studies of HRV circulation and type distribution over a wide geographical area, it is difficult to speculate on the significance of these seemingly rare types. It is possible that these represent types which have a tendency to cause more mild and self-limiting infections, which would not be seen on screening carried out among hospitalised patients. No single HRV type with a large number of VP4/VP2 sequences available was found to display geographical restriction, which suggests that most HRV types circulate freely worldwide. A generalised increase in the practice of routine HRV typing will allow further investigation of the geographical distribution of HRV types.

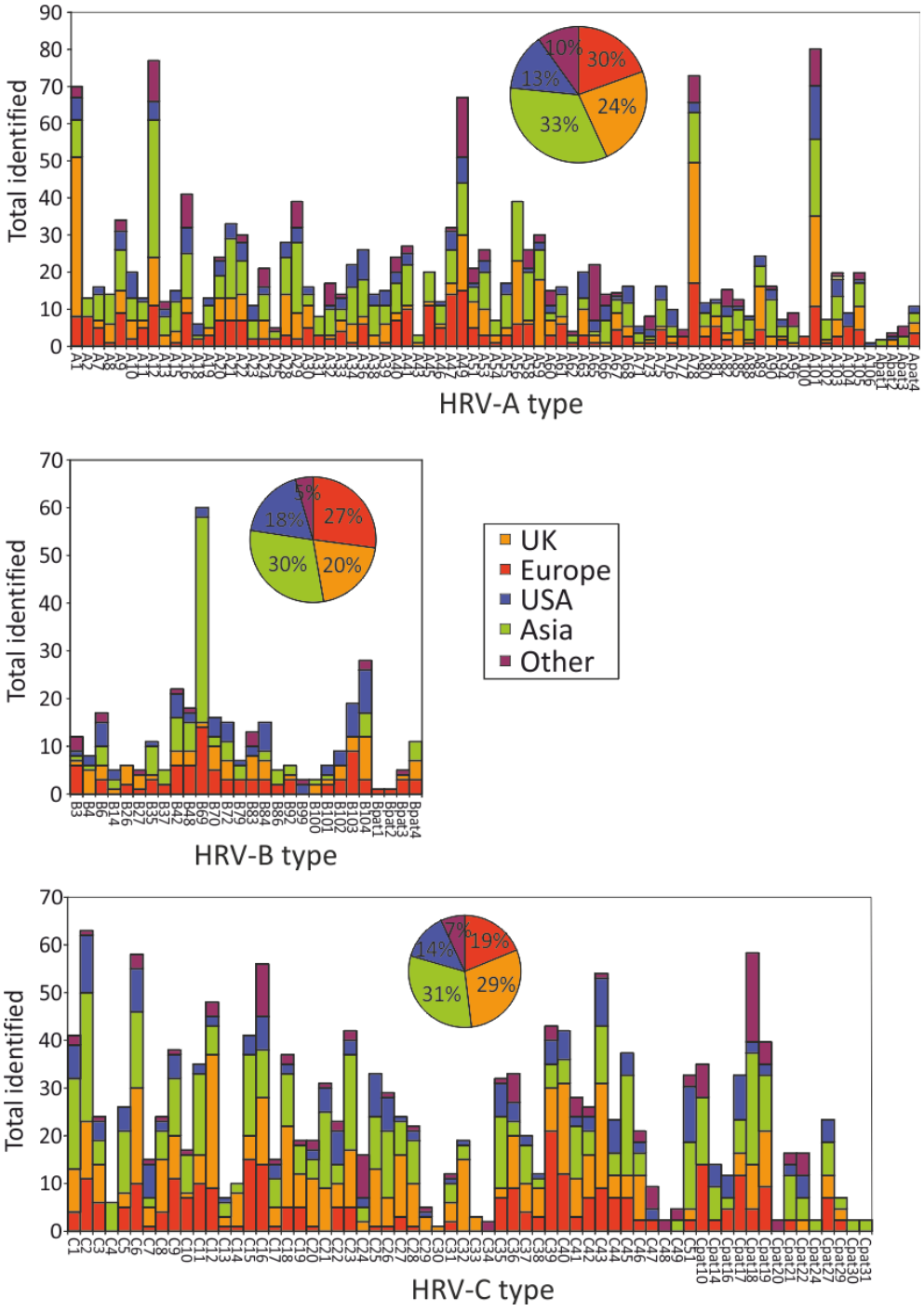


FIGURE 4.8: Total number of identified HRV strains divided by geographical location of isolation. *HRV-A, -B and –C are shown separately. The total representation of sequences from each geographical area is shown in inset pie-chart.*

4.4 Discussion

4.4.1 Development of a sensitive and specific RT-PCR for the amplification of the VP1 region of HRV

The ability to perform genotyping directly from clinical specimens removes the need for cell culture passage and therefore greatly reduces time required for an accurate diagnosis and exposure of staff to live cultures. This has been achieved in studies of EV (Mirand et al., 2009; Nix et al., 2006; McWilliam Leitch et al., 2009b). The HRV specific assays developed in the course of this work allow the amplification of the full VP1 region from respiratory specimens, which often have low viral titres. Unfortunately, due to the high degree of genetic diversity, it was not possible to develop an assay capable of amplifying all three species of HRV concurrently. Indeed, even within a single HRV species, maintaining the required balance between allowing sufficient primer degeneracy to amplify all types and avoiding non-specific amplification was challenging.

Several studies have advocated the use of a one-step PCR for EV typing, comprising only of a first round of amplification (She et al., 2010; Kiang et al., 2009), with the suggestion that a "closed system" is less likely to suffer from contamination. However, one of the above studies reported that almost one third of tested isolates could not be effectively amplified in the VP1 region and those that were successful showed significant non-specific product (She et al., 2010). We have found that a nested PCR method dramatically reduces non-specific amplification and is much more effective for the amplification of very low titre samples. Also, with effective and universal adherence to the "one-way" system in PCR laboratories, with spatial separation of dedicated areas for nucleic acid extraction, first round and second round PCR, the occurrence of contamination is greatly reduced (Chapter 2).

4.4.2 The importance of genotyping in HRV

The identification and definition of HRV types by analysis of available sequence data within the capsid region has the potential to revolutionise HRV typing. The ease with which this can be carried out, requiring very little additional equipment or training, should make the endeavour of classifying all detected HRV sequences much more feasible and attractive. A global system of type identification should allow large-scale investigations of epidemiology, transmission and evolution. Classification and centralised reporting of new HRV types improves the ability of researchers to commence study of the distribution of new types quickly. Whereas previously a considerable latent period would have been necessary while waiting for type-specific antiserum to be developed and dispatched, now sequence data for comparison is available worldwide almost instantly.

4.4.3 Proposed criteria for the division of HRV into genotypically defined types

The usage of nucleotide divergence thresholds for the definition of new EV types for several years with great success strengthens the assertion that this is a valid method of classification for the closely related HRV. The criteria that we have developed and have been proposed formally (Simmonds et al., 2010; McIntyre et al., 2013a) are summarised below. The overall conclusions and recommendations are informed by and consistent with those already used for EV. However, by incorporating lower thresholds and acknowledging the lack of capsid recombination by proposing a VP4/VP2 divergence threshold, we have taken into account the differences between HRV and EV in terms of biological properties and genetic diversity.

Proposed criteria for the assignment of genotypically defined types of HRV

1. An HRV type should be phylogenetically distinct from other HRV types.
2. Analysis of divergence in the VP1 region should include at least 90% of the full VP1 coding region for each species.
3. In the VP1 region, an HRV type should have at least 13% (HRV-A), 12% (HRV-B) or 13% (HRV-C) nucleotide divergence from all other HRV types.
4. In cases where nucleotide divergence in VP1 is not sufficient to definitively assign types, phylogenetic relationships may be taken into account.
5. VP4/VP2 sequences can be used for type identification in epidemiological studies.
6. If a putative new HRV type is initially identified from VP4/VP2 sequences and no VP1 sequence data is available, it should be designated as a *provisionally assigned type.*
7. For putative new types that are confirmed by VP1 nucleotide divergence, a permanent HRV type designation should be assigned.
8. New HRV type names should be numbered sequentially and include the species designation (eg/ HRV-A104).
9. The "prototype strain" of each new HRV type should be the first reported full genome sequence or, if this is unavailable, the first reported full VP1 sequence.
10. The Picornavirus Study Group should continue to oversee assignment of new HRV types, as the group includes members with specific expertise in enterovirus type assignment and a number of scientists currently active in HRV research.

The use of only capsid coding regions in type assignment criteria should not detract from the importance of continued investigation of other genomic regions, in particular where these may contribute to the phenotype and disease associations of HRV. In addition, the use of the VP1 region for definition of new HRV types should not discourage the widespread continued use of 5ˈUTR and VP4/VP2 screening protocols. Screening with these relatively conserved regions allows a much greater opportunity for discovery of previously unknown HRV types.

Although this represents, to our knowledge, the most comprehensive survey of HRV sequence data currently possible, the guidelines should be subject to further review as additional data becomes available.

4.4.4 Future directions

In all three species, the phylogeny of the full VP1 region reliably separated isolates of known differing type into distinct bootstrap supported clades. As VP1 also contains a large number of the neutralizing immunogenic sites observed in HRV-A and -B (Kistler et al., 2007b), we can assume that this region is an accurate molecular determinant of the serotype. An interesting potential future addition to this work would be confirmation of serological properties of HRV-C type groups. This would require an easily implementable system for the cultivation of HRV-C *in vitro*. The sinus mucosal organ culture system currently available represents an excellent opportunity to study biological properties of HRV-C (Bochkov et al., 2011). However, it is not suitable for the extensive use that would be required to confirm type designations as not only is it more labour-intensive than other cell culture methods, but requires regular access to a local Department of Head and Neck Surgery to obtain viable sinus tissue. Limited serotyping studies would also be useful in re-confirming some potentially aberrant HRV-A and HRV-B type assignments where cross-reactivity is uncertain (such as HRV-A54/A98).

There is a possibility that a smaller fragment of VP1 could be suitable for genotyping in HRV. Oberste et al highlighted that a 450 base segment at the 3' end of VP1 could be used for EV typing and the results had a 100% correlation with neutralization results (Oberste et al., 1999a). A second study showed full correlation of a 303 nucleotide stretch of VP1 with neutralization data, but this was only applied to 59 Genbank strains (Kiang et al., 2009). In addition, a simple, fast and effective method of typing EV by pyrosequencing has recently been developed (Silva et al., 2008). This may become extremely useful as pyrosequencing becomes more widely available worldwide.

Finally, there is potential for the development of an interactive web tool which would allow users to input a FASTA sequence and receive an HRV type designation. The

service could include an additional step comprising of request for input of a VP1 sequence and subsequent analysis in cases where type designation is dubious. A similar system is already in use for EV and Noroviruses (Kroneman et al., 2011). However, this would have to be modified significantly to allow assignment of HRV types, as the described algorithm is based solely on phylogeny. In order to accurately assign HRV types, it would be desirable to consider both inferred phylogeny and calculated pairwise nucleotide p-distances from all prototype strains. Such a tool, widely and free available on the internet, would make HRV genotyping even more freely accessible, easy and attractive to research groups the world over.

Chapter 5
Analysis of genetic diversity and sites of recombination in HRV-C

5.1 Introduction

5.1.1 Recombination in the evolution of human enteroviruses

Human EV are divided into four species by phylogenetic analysis (EV-A-D) and the three serotypes of PV are now classified as EV-C. EV-B is the largest and most clinically common human EV species. Developing a detailed understanding of the frequency and occurrence of recombination in circulating EV has been vital in surveillance of outbreaks of vaccine derived poliomyelitis, whereby vaccine PV strains may recombine either with each other or with circulating EV-C strains to produce neurovirulent PV (Cuervo et al., 2001; Blomqvist, 2003; Jegouic et al., 2009).

Circulating EV strains undergo frequent recombination, which manifests as observed incongruities between inferred phylogenies in distant genome fragments. Intra-species recombination within the coding region has been described in EV-A (Simmonds and Welch, 2006; Huang et al., 2008; McWilliam Leitch et al., 2012; Santti et al., 1999), EV-B (Simmonds and Welch, 2006; Lukashev et al., 2005; Lindberg et al., 2003; McWilliam Leitch et al., 2009a, 2010; Chevaliez et al., 2004; Andersson et al., 2002) and EV-C (Jegouic et al., 2009; Dahourou et al., 2002). EV has a similar genome structure to HRV with four structural and seven non-structural proteins translated as a single polyprotein that is cleaved post-translationally. Recombination breakpoints are often observed to occur within the non-structural protein coding region and frequently near the boundary with the structural protein coding region. Specifically, recombination breakpoints have been isolated in the 2A

protease coding region (van der Sanden et al., 2011), 2B (Lindberg et al., 2003; Chevaliez et al., 2004), 2C (Lukashev et al., 2005; Andersson et al., 2002), at the 5' end of 3D (van der Sanden et al., 2011) and (in PV recombinants generated *in vitro*) within the 3C region (Runckel et al., 2013).

The frequent recombination between the capsid and non-structural protein coding regions has led to a theory that these genome regions of EV strains evolve in a modular and independent fashion (Santti et al., 1999; Lukashev et al., 2005). This implies that EV diversity is derived from constant recombination between capsid sequences and pools of non-structural genome sequences. The combination of a capsid sequence and non-structural region generates a specific recombinant form (RF). These RFs have been shown to rise to dominance and then disappear entirely over the span of only a few years (McWilliam Leitch et al., 2009a).

In contrast, only rarely is recombination observed within the capsid coding region of EV (Santti et al., 1999; Lindberg et al., 2003; McWilliam Leitch et al., 2009a; Simmonds and Welch, 2006) and this has been previously attributed to the significantly greater sequence divergence seen within the capsid coding region (Lindberg et al., 2003). It has been suggested that the increased divergence may present a biological compatibility barrier, which renders variants with chimaeric capsid regions unviable or simply less fit than parental strains (Simmonds and Welch, 2006). The theory of the biological compatibility barrier is additionally supported by the observation that inter-species recombination in any part of the coding region is only very rarely observed in EV. One recent study of PV recombination *in vitro* used a deep sequencing technique to isolate not only viable PV recombinants, but all genomes generated during viral replication (Runckel et al., 2013). Analysis of recombination breakpoints in over 50,000 generated genomes found that there was actually no difference between the frequency of occurrence of recombination within the capsid and the non-structural protein coding region. This

supports the assertion that capsid recombinants are likely routinely randomly generated, but are less able to compete with wild type viruses. Viable circulating EV strains with chimaeric capsids have been infrequently isolated (Blomqvist, 2003; Bouslama et al., 2007; Zhang et al., 2010). In addition, one study found a recombination breakpoint near the 5ˈ end of the VP4 region, presumably related to recombination within the 5ˈUTR (Lukashev et al., 2005).

Within all four species of human EV, 5ˈUTR sequences cluster phylogenetically into two large groups (Pöyry et al., 1994). Sequences from EV-A and –B are interspersed within one group (Group I) and sequences from EV-C and –D within the other (Group II) (Pöyry et al., 1996). Extensive phylogenetic incongruence has been observed between the 5ˈUTR and capsid coding regions (Simmonds and Welch, 2006) and no evidence has been found of a correlation between phylogenetic grouping in the 5ˈUTR region and other putative recombinant regions downstream (Lukashev et al., 2005). In addition, four recently described new EV types (EV-A90, EV-A91, EV-C104 and EV-C109) have 5ˈUTR sequences which group outwith the 5ˈUTR cluster expected by their EV species designation (Yozwiak et al., 2010; Tapparel et al., 2009b; Smura et al., 2007a). Two of these are EV-C types (EV-C104 and EV-C109) which group closely with EV-A/-B sequences in the 5ˈUTR and two (EV-A90 and EV-A91) are EV-A types which cluster with EV-C/-D in the 5ˈUTR. This indicates a degree of reciprocal exchange between the two 5ˈUTR clusters in the contemporary evolution of EV strains.

5.1.2 Recombination in the 5ˈUTR of HRV-C

HRV-C was only discovered relatively recently (McErlean et al., 2008; Arden et al., 2006; Kistler et al., 2007b; Lamson et al., 2006; Renwick et al., 2007; Lau et al., 2007) and at the time that this work was undertaken, only limited information regarding the frequency and occurrence of recombination within the HRV-C genome had been published (Huang et al., 2009; Wisdom et al., 2009a). However,

recombination was thought to be rare within the coding region of HRV-A and -B (Kistler et al., 2007b; Lewis-Rogers et al., 2009). In contrast, preliminary studies undertaken of circulating genetic diversity of HRV-C strains indicated that they fell into two distinct phylogenetic groups (Han et al., 2009; Huang et al., 2009; Wisdom et al., 2009a). In fact, over 60% of sequenced strains which are members of HRV-C species by sequence analysis in the capsid region group within the HRV-A clade in the 5'UTR (Wisdom et al., 2009a). It appears that similar to EV strains, inter-species recombination in the 5'UTR has taken place extensively in the evolutionary history of this species. Studies of the clinical correlates and molecular epidemiology of HRV-C strains have found no difference in the clinical outcomes or circulation patterns of recombinant (HRV-Ca) or non-recombinant (HRV-Cc) strains (Huang et al., 2009; Wisdom et al., 2009a).

The grouping of the majority of HRV-C strains within the HRV-A clade has probably contributed to the relatively recent discovery of HRV-C, despite the fact that it appears to have been circulating in human populations for many years (Briese et al., 2008). Prior to the recognition of recombinant HRV strains, it was common practice to use 5'UTR sequence comparisons for HRV species and type identification (Kiang et al., 2008); a practice which would have led to the misidentification of over two thirds of HRV-C strains as HRV-A. However, it is now generally accepted that this typing method is fundamentally flawed (Savolainen-Kopra et al., 2009b) and the majority of published studies utilise the VP4/VP2 region instead.

5.1.3 Genetic features of HRV-C

HRV-C strains share the same basic genome organisation with HRV-A and -B and other EVs, including an approximately 7kB genome encoding a single open reading frame. HRV-C have been found to have a generally shorter coding region than other *Picornaviridae*, secondary to deletions within the VP1 capsid protein coding region (Lau et al., 2007; Arden et al., 2010). Also in common with other HRV are the

observed amino acid sequence motifs at the start of the coding region (MGAQVS) and within the 3D polymerase coding region (YGDD, YGL, FLKR, SIRWT) (Lau et al., 2007; Arden et al., 2010; McErlean et al., 2008). These are all fully conserved in other HRV species (Savolainen et al., 2004) and the 3Dpol motifs are additionally conserved in EV strains (Brown et al., 2003).

Distinguishing features of HRV-C genomes include the unique position of the *cis-acting* replication element (cre) within VP2 (Cordey et al., 2008; Arden et al., 2010). In contrast, this has been located within 2A in HRV-A (Gerber et al., 2001) and VP1 in HRV-B (McKnight and Lemon, 1998). In addition, HRV-C are found to have a generally elevated G+C content of between 42-43% compared to HRV-A and –B (39-40%), although this is generally lower than the G+C content seen in EV (46%) (Arden et al., 2010; Lau et al., 2007).

Similar to all members of the genus *Enterovirus*, HRV-C strains possess a 5ˈUTR of approximately 600 bases. The 5ˈUTR is a highly structured and well conserved region which contains certain elements essential to effective replication of the virus. As all EV, HRV possess a type I IRES, which is essential for the internal initiation of translation of the uncapped mRNA (Palmenberg et al., 2009). In addition, a 5ˈ cloverleaf structure, implicated in the initiation of RNA synthesis was present and structurally conserved in all three species of HRV (Palmenberg et al., 2009). The high degree of conservation of both RNA sequence and secondary structure observed within this region may contribute to the observed propensity for recombination within the 5ˈUTR of HRV-A/-C.

The majority of the analysis presented within this chapter focussed on the occurrence of recombination within samples collected from Edinburgh patients between 2006 and 2007. This has been previously published (McIntyre et al., 2010). However, section 5.3.2 also refers to an additional analysis of the 5ˈUTR region of all

published HRV-C full genomes, undertaken in 2012 and published as part of a further paper (McIntyre et al., 2013b).

5.2 Materials and Methods

5.2.1 Sample selection

A previously conducted study had produced sequences from VP4/VP2 and a small fragment of the 5'UTR of 144 HRV-C sequences derived from Edinburgh patients. These fell into 46 phylogenetic clusters, now known to be analogous to type groups. Samples for inclusion in the study of recombination within the coding region of HRV-C were selected in order to include at least one representative from as many of these phylogenetic groups as possible. In addition, a further 15 samples from epidemiological studies (Chapter 3) were selected. This resulted in the inclusion of 35 HRV-C types in the 5'UTR, VP4/VP2 and 3Dpol regions and 28 in the VP1 region.

5.2.2 Amplification of the VP4/VP2, VP1 and 3Dpol regions of HRV-C

The VP4/VP2 and VP1 regions were amplified as previously described (Chapter 3, Chapter 4). A 3Dpol fragment (from positions 6384 – 6854 numbered by EF582385) was amplified using a nested PCR protocol (described in Chapter 2 and primers listed in Appendix 1). Samples negative on initial nested PCR were amplified using a SuperScript III protocol (Chapter 2).

5.2.3 Amplification of the 5'UTR region of HRV-C

Two sequence fragments with a 61 base overlap were produced, allowing a composite 5'UTR sequence of 680 bases to be constructed. The amplification of the 5'UTR region utilised the two primer sets (UTR2 and UTR3). In addition, a further 12 HRV-C variants were further amplified using hemi-nested combined HRV-A/HRV-C primers to give sequence complete from nucleotide position 27 to 355

(primer set UTR1 in Appendix 1). Cycling conditions for all reactions are described in Chapter 2.

5.2.4 Amplification of the 2A region of HRV-C

The 2A coding region of HRV-C strains was amplified in two fragments with a 303 base overlap, to produce a single sequence of 884 bases. Both second round PCRs used the same first round PCR product. Due to the fact that the first round fragment to be amplified was almost 1000 bases in length, we found that the addition of 2µL 3mM dNTPs in the first round gave a higher sensitivity (concentration of nuclease-free H_2O was adjusted accordingly to give a final reaction volume of 20µL). Resulting amplified DNA bands of the appropriate size were then cut out from agarose gels visualized under UV transillumination. The target DNA was then extracted from the agarose gel and sequenced as previously described (Chapter 2).

5.2.5 Dataset construction

A total of 89 HRV-C positive samples corresponding to 40 phylogenetic groups were amplified in the VP4/VP2 and 3Dpol regions[14]. Two samples (Resp_3856 and Resp_6131) could not be amplified in the 5'UTR, possibly due to low viral loads in the original sample. However, both of these samples came from phylogenetic groups in which other members were represented within the 5'UTR dataset.

Of the 89 HRV-C positive samples with sequences available for the 5'UTR, VP4/VP2 and 3Dpol regions, only 71 could be successfully amplified in the VP1 region. This corresponded to 33 of the 40 phylogenetic groups. This was due to limitations in the sensitivity of the originally described HRV-C VP1 nested PCR and

[14] 74 of the included VP4/VP2 sequences had been previously amplified as part of a previously published study (Wisdom et al., 2009a).

subsequent redevelopment of this protocol led to a much improved sensitivity[15] (McIntyre et al., 2013b). Sequences which could not be amplified in the VP1 region did not cluster together on phylogenetic trees in other regions. We can therefore conclude that these did not represent an identifiable genetic subset whose omission would have significantly affected the analysis.

Sequences generated in the course of this study have been submitted to Genbank and assigned accession numbers as described (Table 5.1). In addition, VP4/VP2 sequences generated as part of the previous study by Wisdom et al are published with accession numbers in the series GU294336 to GU294480 (Wisdom et al., 2009a).

TABLE 5.1: Accession numbers of sequences generated during the study of recombination in HRV-C in 2006/2007

Genome Region	**Number of sequences**	**Accession Numbers**
VP4/VP2	15	HM352737 – HM352752
VP1	71	HM236897 – HM236968
3Dpol	89	HM485468 – HM485556
5'UTR	86	HM581802 – HM581888

All available HRV-C full genomes (n=10) were initially obtained from Genbank in April 2010. For the inter-species recombination and full genome sequence divergence analysis, these were combined into a single dataset with available HRV-A (n=82) and HRV-B (n=25) full genome sequences. An additional analysis of the 5'UTR region of all available full genome sequences of HRV-A and HRV-C was undertaken in July 2012. This included all full genome sequences which had a 5'UTR fragment that was more than 90% complete across the region 167-626 (numbered by FJ445111 : HRV-A1). 167 HRV-A and 59 HRV-C sequences were included in the analysis.

[15] Recombination studies which utilised this improved HRV-C VP1 nested PCR protocol are detailed in Chapter 6. In addition, HRV-C36-C46 VP1 sequences were generated using the improved HRV-C VP1 PCR.

Phylogenetic trees were constructed as previously described (Chapter 2) and recombination analysis was undertaken primarily by inspection of phylogenetic trees for bootstrap supported changes in tree topology or branch length. This analysis was supplemented by screening datasets with the TreeOrderScan software within the SSE v1.0 package (Simmonds, 2012). Breakpoints were determined by analysis with the program GroupScan within the SSE v1.0 software package.

5.3 Results

5.3.1 Phylogenetic analysis of the VP4/VP2, VP1 and 3Dpol regions of HRV-C

For analysis of recombination within the coding region of HRV-C, three genome regions (VP4/VP2, VP1 and 3Dpol) were included. VP1 sequences are the basis for type designations of HRV-C (Simmonds et al., 2010) whereas VP4/VP2 and 3Dpol represent the extreme 5' and 3' ends of the coding genome. In all three regions, sequences fell into clearly defined, bootstrap supported groups (Figure 5.1). A total of 15 HRV types were represented by a single sequence only and the remainder were represented by several unique variants. In both VP4/VP2 and 3Dpol, the majority of sequences clustered according to their type assignment, as defined by VP1. The phylogeny of all three coding regions was remarkably congruent, in terms of both clade membership and branch length.

Only one HRV-C sequence showed evidence of incongruent phylogenetic position (shaded box in Figure 5.1). GQ223227, a previously published full genome sequence (Huang et al., 2009), is the HRV-C8 prototype strain as defined by VP1 (Simmonds et al., 2010). Our sample selection contained one other HRV-C8 variant (Resp_10300) confirmed by phylogenetic grouping and pairwise nucleotide p-distances. However, while Resp_10300 maintained its phylogenetic position relative

to other sequences in 3Dpol region, GQ223227 changes tree position to group with the HRV-C9 prototype strain, GQ223228 in 3Dpol.

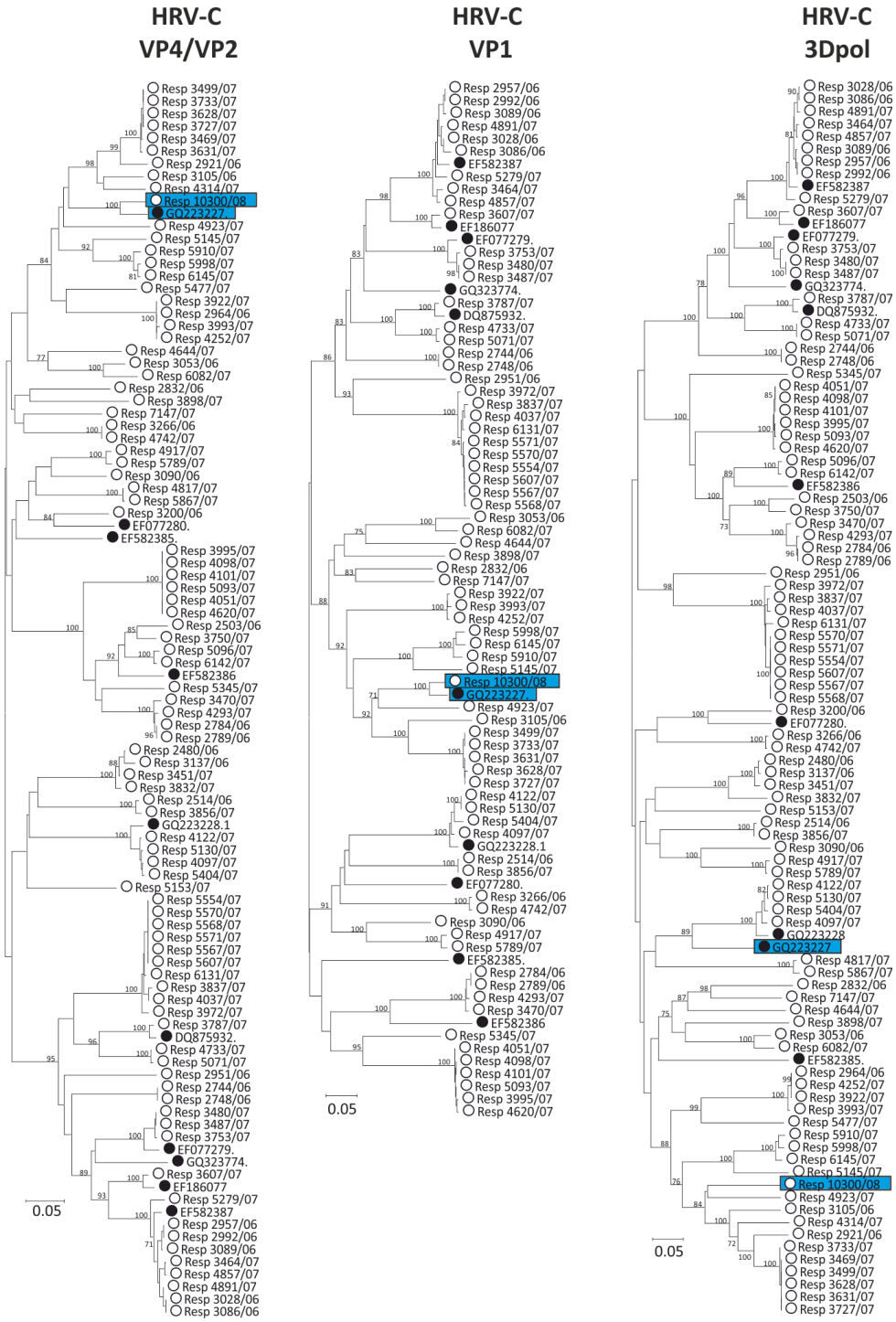


FIGURE 5.1: Neighbour joining phylogenetic trees of study sequences and corresponding regions from published complete genome sequences from the VP4/VP2, VP1 and 3Dpol regions of HRV-C. *Branches showing at least 70% bootstrap support are indicated. Branch to tree root has been collapsed for ease of reference. Complete genome sequences were labelled in black. The putative recombinant sequence group containing GQ223227 and Resp_10300 (HRV-C8) are marked with blue boxes. Branches are scaled by genetic distance.*

On inspection of these two full genome sequences, the change in tree topology appeared to occur within 3Dpol. At the 5' end of 3Dpol (position 6384 – 6645), GQ223227 had a pairwise nucleotide p-distance of 0.08 from Resp_10300. However, at the 3' end of 3Dpol (position 6646 – 6854), this sequence was nearly identical (p-distance 0.01) to GQ223228 and was much more distantly related (p-distance 0.35) to Resp_10300. Inspection of phylogenetic trees for the 5' and 3' ends of 3Dpol clearly showed the change in grouping of GQ223227 within the region (Figure 5.2).

This observation is complicated by the fact that both GQ223227 and GQ223228 originated from the same published study within the same laboratory (Huang et al., 2009). Although we cannot definitively exclude that this sequence represents a genuine natural recombination event, in light of the extremely rare nature of recombination within the rest of the sample set and the origination of both segments of this chimaeric genome from the same laboratory, it seems prudent to question its authenticity. As the other HRV-C8 sequence available within our dataset shows no evidence of this putative recombination event, it was felt that the hybrid genome may instead represent a sequencing or assembly error. As a result, GQ223227 was not considered as a genuine recombinant genome in the remainder of the analysis carried out.

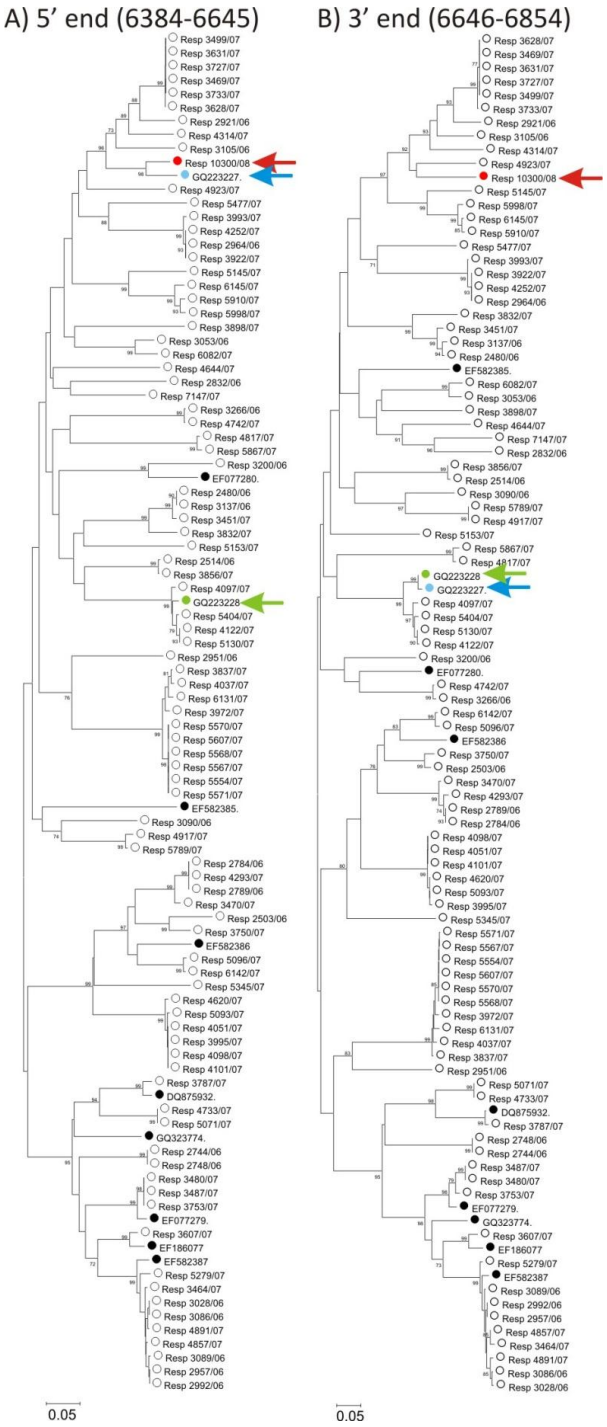


FIGURE 5.2: Comparison of phylogenies obtained from analysis of the 5' and 3' ends of 3Dpol of HRV-C sequences. *Published full genomes are marked with black circles. Sequences involved in putative recombination event are labelled with coloured arrows.*

5.3.2 Phylogenetic analysis of the 5'UTR of HRV-C

Phylogenetic trees constructed for the 5'UTR region of HRV-C sequences from 2006-2007 (and including 10 HRV-C full genomes) showed that sequences continued to group within their types (as defined by VP1) in the 5'UTR. Even the putative recombinant HRV-C8 type (consisting of GQ223227 and Resp_10300) was congruent in the 5'UTR (marked by shaded grey boxes on Figure 5.3). However, there was extensive breakdown in the deeper relationships between HRV-C type groups observed between the 5'UTR and VP4/VP2 regions.

In accordance with previous reports (Wisdom et al., 2009a; Huang et al., 2009), we found that HRV-C sequences grouped into two distinct clades in the 5'UTR. The largest of these clades (marked as HRV-Ca on Figure 5.3) contains HRV-C sequences which group within the HRV-A clade[16]. In total, over 70% of analysed HRV-C sequences (n=68) representing 30 of the 40 included HRV-C types grouped with HRV-A sequences. The remaining 10 HRV-C types grouped in a distinct clade, labelled as non-recombinant HRV-Cc strains (marked in red on Figure 5.3). Putative recombinant (HRV-Ca) strains were scattered within the VP4/VP2 region, indicating the possibility that recombination between the 5'UTR of HRV-A/-C has occurred on multiple occasions throughout HRV-C diversification (shown by blue lines and grey arrows on Figure 5.3).

Analysis of 5'UTR and VP4/VP2 sequences with TreeOrderScan showed a minimum of 12 phylogeny violations between the two regions (data not shown). On inspection of phylogenetic trees, inter-type relationships showed a great deal of inconsistency between the two regions, with some large clades being entirely maintained and others showing breakdown in phylogenetic relationships. Two large HRV-Ca clades containing 6 and 8 HRV-C types respectively were bootstrap

[16] For ease of reference, relationships between HRV-C sequences only are presented in Figure 5.3.

supported in both regions with completely congruent tree topology (marked by grey boxes on Figure 5.3).

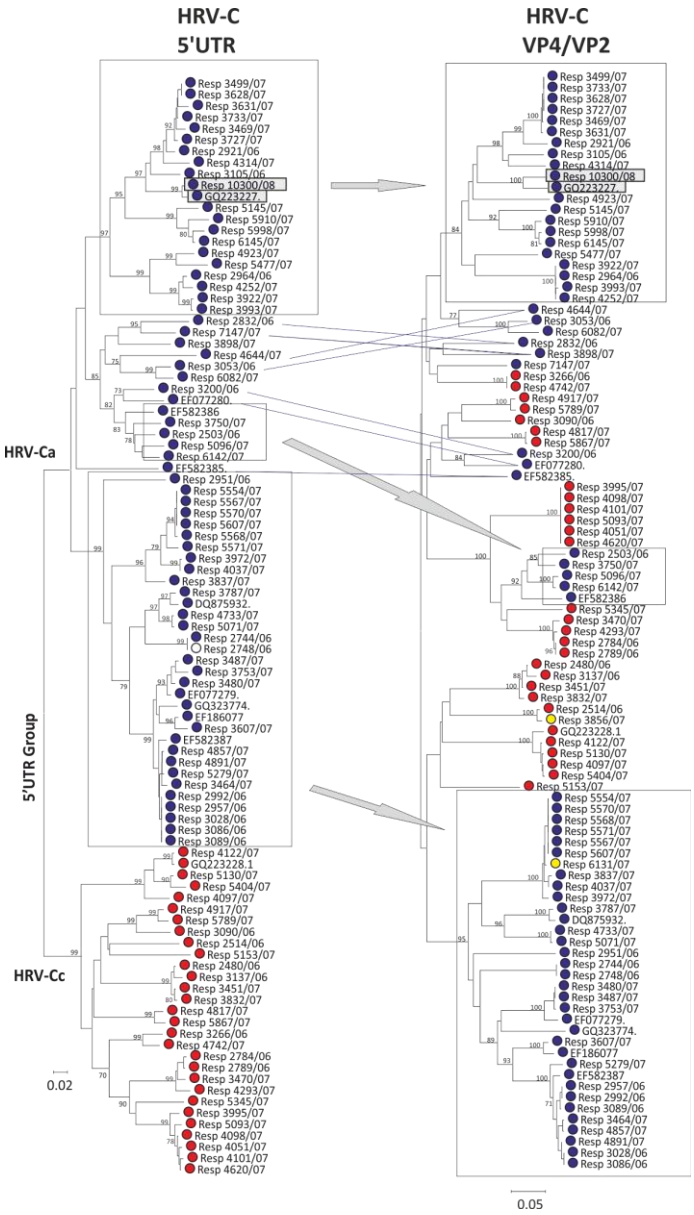


FIGURE 5.3: Neighbour joining phylogenetic trees showing the partial 5'UTR and VP4/VP2 regions. *Putative recombinant (HRV-Ca) sequences are coloured in blue. Non-recombinant (HRV-Cc) sequences are coloured in red. Two samples which could not be amplified in the 5'UTR region are shown in yellow. The phylogenetic position of recombinant HRV-Ca sequences are shown by blue lines. Two contiguous and congruent clades of HRV-Ca sequences are shown by grey boxes and grey arrows. Branches are scaled by genetic distance.*

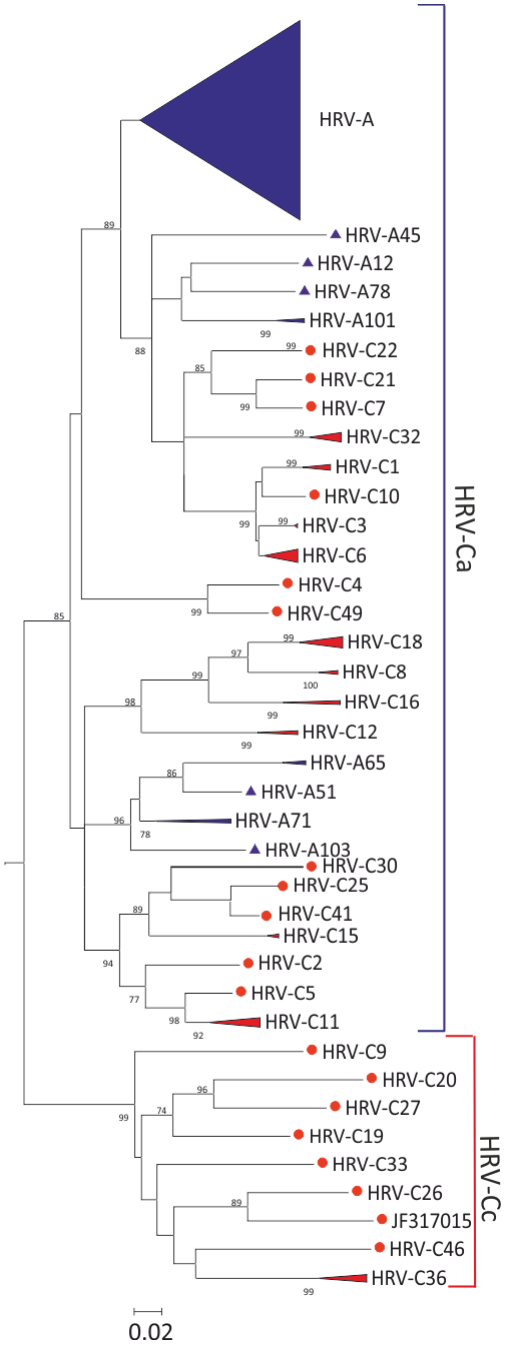


FIGURE 5.4: Neighbour joining phylogenetic tree showing the 5'UTR region of all HRV-A and HRV-C full genomes. *HRV-A sequences are marked in blue, with HRV-C marked in red. Branch to tree root has been collapsed for ease of reference. The largest portion of the HRV-A clade, which had no HRV-C sequences embedded within it has been additionally collapsed. Branches are scaled by genetic distance.*

An additional analysis of all available HRV-A and HRV-C 5'UTR regions from published polyprotein sequences revealed a similar pattern, whereby the majority of HRV-C sequences and type groups were embedded within the HRV-A clade (Figure 5.4). This analysis included 59 HRV-C 5'UTR sequences representing 30 HRV-C type groups. All sequences analysed grouped in the 5'UTR region according to their assigned type. Recombinant sequences were not distributed throughout the entire HRV-A 5'UTR clade, but fell into three distinct groups. One group of HRV-C sequences appeared to share a common ancestor with HRV-A12, -A45, A78 and –A101, while another was closest to HRV-A51, -A65, -A71 and –A103. The third small clade consisted only of HRV-C4 and –C49. Although this group fell within the HRV-A clade, they did not group closely with any known HRV-A sequence.

5.3.3 Determination of recombination breakpoints within the 5'UTR of HRV-C

In order to determine the location of putative recombination breakpoints in the HRV-C 5'UTR, amplified 5'UTR and VP4/VP2 sequences were joined to give a single, long sequence which was scanned between positions 185 – 903 (numbered by EF582385) using the program GroupScan. The sequences generated in the current study were combined with the 10 available published HRV-C complete genomes, 34 5'UTR sequences previously published (Huang et al., 2009) and all available HRV-A/-B complete genome sequences to construct the dataset for analysis. Each putative recombinant (HRV-Ca) sequence was scanned against HRV-A, HRV-B and control HRV-Cc (n=29) sequences and the recombination breakpoint was calculated from the intersection of the two lines (Figure 5.5).

Of the 51 confirmed HRV-C types, 42 had UTR sequences that were more than 90% complete across the region analysed. Of these, 31 HRV-C types were members of the HRV-Ca group and breakpoints were calculated. Calculated recombination breakpoints were similar within HRV-C type groups and all calculated breakpoints

occurred within the UTR (Table 5.2). The majority of HRV-Ca sequences showed putative recombination breakpoints that fell within two "hotspot" regions. These mapped within the IRES to the polypyrimidine tract (PPT) between stem loops 5 and 6 (positions 561-576; mean, 565) and near the terminal loop of stem-loop 5 (position 508-544; mean,523). Only two HRV-C types had recombination breakpoints out with these two regions and both were at a more 5' site (position 479-489; mean, 481).

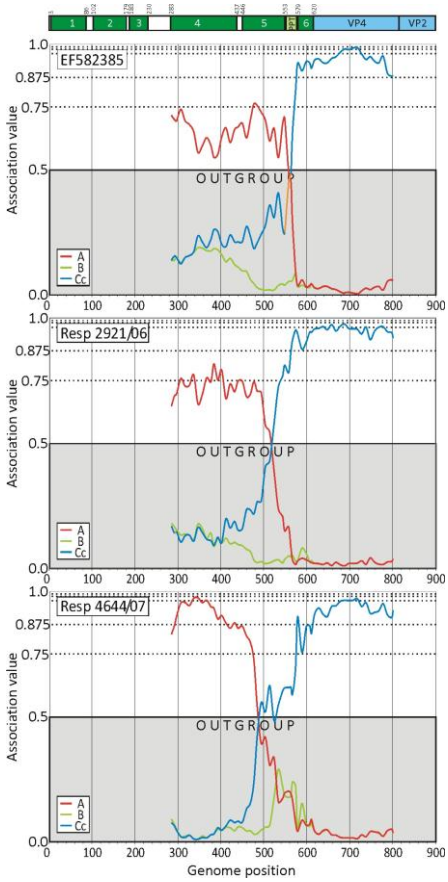


FIGURE 5.5: Recombination breakpoints for three representative HRV-Ca sequences*. Analysis was carried out in the program GroupScan using HRV-A, -B and –Cc sequences as control groups. Above the graphs is a diagrammatic summary of 5'UTR RNA structure elements (stem-loops 1-6* (Anderson et al., 2007) *and polypyrimidine tract [PPT]). Breakpoint is calculated from intersection of two lines. Association values/grouping scores of less than 0.5 refer to distantly related outgroups (marked in blue).*

TABLE 5.2: Calculated 5'UTR recombination breakpoints for HRV-C types analysed

HRV Type	5'UTR group	Total sequences analysed	Breakpoint (nt) mean	Breakpoint (nt) range
C1	Ca	5	520	519-523
C2	Ca	1	564	-
C3	Ca	2	526	526-527
C4	Ca	1	561	-
C5	Ca	3	563	563-564
C6	Ca	10	524	524-526
C7	Ca	2	574	574
C8	Ca	4	565	561-570
C9	Cc	-	-	-
C10	Ca	1	525	-
C11	Ca	2	554	552-557
C12	Ca	4	508	508
C13	Ca	1	523	-
C14	Cc	-	-	-
C15	Ca	3	483	479-489
C16	Ca	3	520	518-526
C17	Ca	1	567	-
C18	Ca	8	528	518-544
C19	Cc	-	-	-
C20	Cc	-	-	-
C21	Ca	6	572	570-575
C22	Ca	2	576	576
C23	Ca	3	562	552-570
C24	Ca	1	560	-
C25	Ca	1	558	-
C26	Ca/Cc[a]	-	-	-
C27	Ca/Cc[a]	3	535	522-559
C28	Ca	2	543	536-550
C29	Cc	-	-	-
C30	Ca	1	561	-
C31	Ca	1	528	-
C32	Ca	9	524	523-527
C33	Cc	-	-	-
C35	Cc	-	-	-
C36	Cc	-	-	-
C37	Cc	-	-	-
C38	Ca	2	555	555
C40	Ca	1	560	-
C41	Ca	1	548	-
C42	Ca	1	479	-
C46	Cc	-	-	-
C49	Ca	1	570	-
C51	Cc	-	-	-

[a] HRV-C26/-C27 are labelled as HRV-Ca/Cc groups, as previously published sequences of these types do not all fall within the same 5'UTR group (Huang et al., 2009; Wisdom et al., 2009a).

An analysis of mean pairwise p-distance within and between HRV species groups (with HRV-Ca and HRV-Cc considered separately) confirmed high sequence conservation within the 5'UTR. In particular, sequences of the stem-loop regions either side of the recombination hotspot centring on position 523 were highly conserved (Figure 5.6). A further analysis of this data performed by Professor P. Simmonds showed that HRV-Ca variants with a 5' sequence that was HRV-A like and a 3' sequence that was closely related to HRV-C showed similar predictive pairing to both non-recombinant HRV-A and HRV-Cc sequences. It was therefore hypothesized that this region could comfortably accommodate a recombination event without destabilising the stem-loop and therefore disrupting the function of the highly structured IRES.

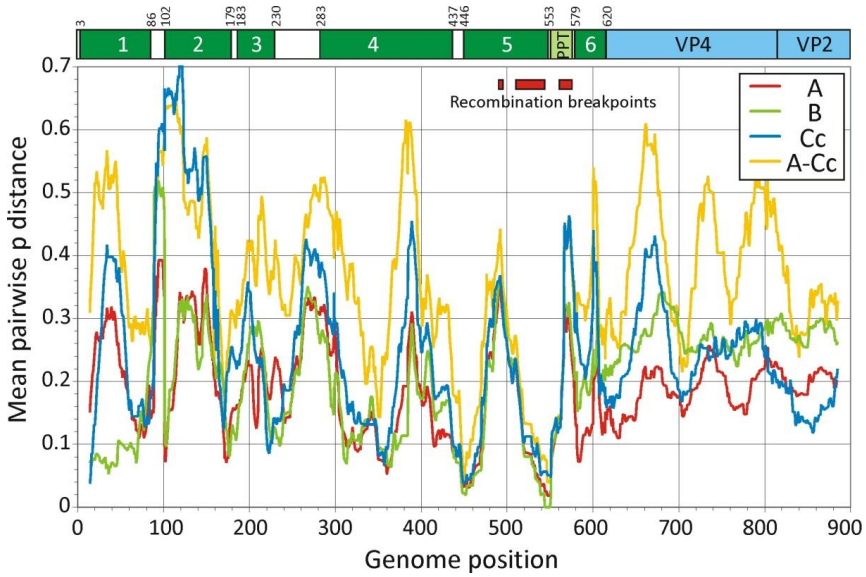


FIGURE 5.6: Sequence variability scan of the 5'UTR showing mean HRV-A, -B and –Cc inter-type distances and interspecies distance of HRV-A/Cc. *The pairwise p-distances are averaged over a 30 base window. The position of known 5'UTR structures and putative location of recombination breakpoints are shown.*

Within the fragment of 5'UTR analysed (from position 185), all sequences showed evidence of only one recombination breakpoint. At the 5' end of the sequence fragment, all sequences grouped strongly within the HRV-A clade. In order to

investigate whether a 5' breakpoint existed in HRV-C 5'UTR sequences, a subset of 11 HRV-Ca variants were sequenced to position 27. All 11 sequences analysed grouped with HRV-A at the extreme 5' end of the genome. In each case, association values with HRV-A at the 5'end were significantly higher than the 0.5 score which signifies an outgroup position (an example is shown in Figure 5.7). Analysis of phylogenetic trees constructed from 50 and 100 base fragments of this region showed no evidence for grouping of any HRV-Ca variants outside the HRV-A clade (data not shown).

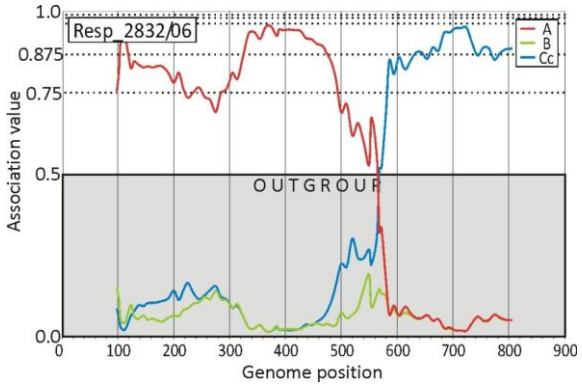


FIGURE 5.7: GroupScan analysis of the extreme 5' end of the HRV-C genome. *This analysis used a subset of control sequences which were complete from position 27 to scan against the 11 HRV-Ca query sequences.*

5.3.4 Sequence divergence across the full genome of HRV-C

Mean pairwise nucleotide and amino acid p-distances were calculated for the VP4/VP2, VP1 and 3Dpol regions of HRV-A, -B and –C (Figure 5.8)[17]. Across all three species, VP1 was uniformly found to be the most divergent region (pairwise nucleotide p-distance of 0.290, 0.265 and 0.347 respectively), while VP4/VP2 was consistently the most conserved region. A significantly greater within-species divergence in all three regions was seen in HRV-C, compared with HRV-A and –B. In fact, the amino acid divergence of HRV-C in both VP4/VP2 and 3Dpol is almost twice that observed in other species.

[17] For HRV-C, only the 10 published full genomes available at the time of this analysis and the sequences generated in this study were included in this calculation.

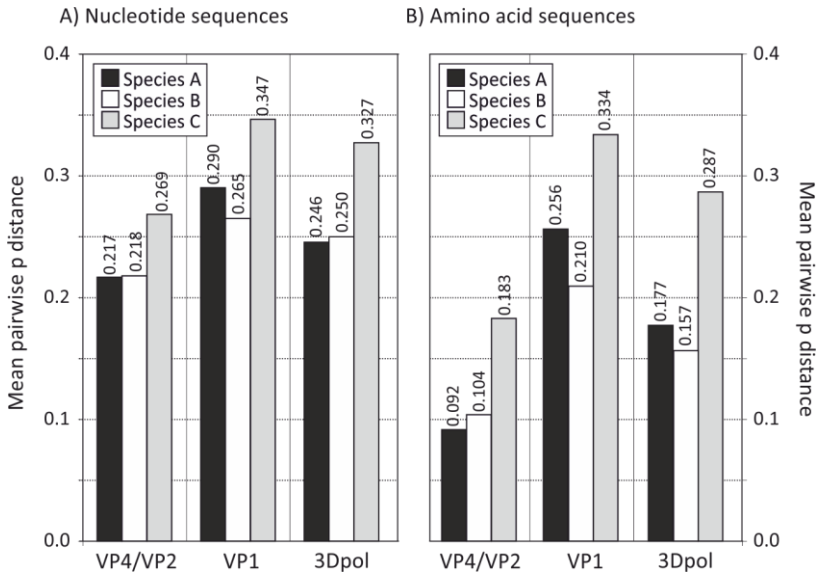


FIGURE 5.8: Mean pairwise (uncorrected) p-distances of nucleotide (A) and amino acid (B) sequences in the VP4/VP2, VP1 and 3Dpol regions of all three HRV species. *Actual values are shown above each bar.*

In order to investigate sequence divergence across the full HRV-C genome and highlight any other potential areas of recombination, a divergence scan using a fragment length of 300 nucleotides, an increment of 30 nucleotides and including all HRV-A/-B and 10 HRV-C full genomes was performed (Figure 5.9). Sequence divergence of HRV-C compared with HRV-A or HRV-B was uniformly higher throughout the genome than within species divergence of HRV-C. This was true for both nucleotide and amino acid p-distances. Interestingly, a small (approximately 260 base) region showed inter-species HRV-A/HRV-C divergence that approached intra-species HRV-C divergence. This area of reduced inter-species divergence fell approximately between positions 3268 and 3525 within the 2A proteinase coding region. HRV-B sequences remained divergent from both HRV-A and HRV-C within this region. This highlighted the possibility of inter-species recombination involving the 2A coding region of HRV-C sequences.

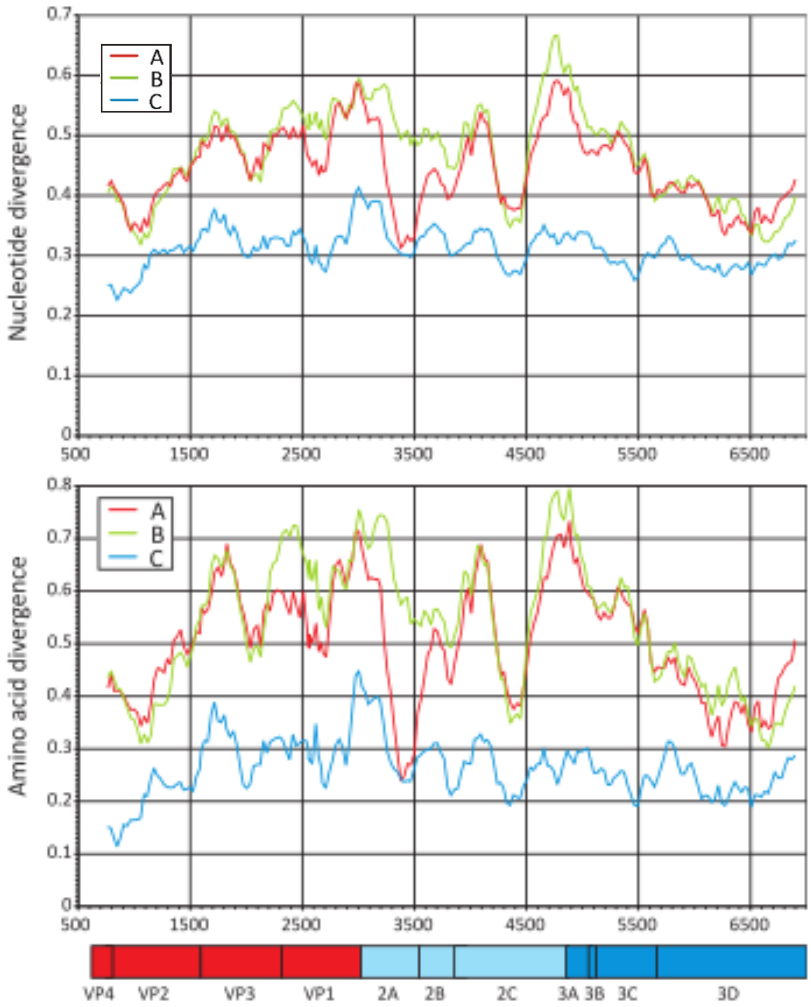


FIGURE 5.9: Sequence divergence scans showing mean nucleotide and amino acid divergences within complete genome sequences of HRV-C and between HRV-C and HRV-A/-B. *Mean inter-species pairwise distances between HRV-A and HRV-C are shown in red, between HRV-B and HRV-C are shown in green and mean intra-species distance for HRV-C is shown in blue. Fragment length of 300 nucleotides and an increment of 30 nucleotides was used. A genome diagram showing different HRV-C genes is shown underneath (with annotation taken from sequence EF582385). The P1 region is marked in red, P2 in pale blue and P3 in blue.*

5.3.5 Analysis of recombination within the 2A region of HRV-C

Initially, in order to further investigate the approximately 260 base region of decreased divergence between HRV-A and HRV-C, phylogenetic trees were constructed of the putative recombinant region and 5'/3' flanking regions of equal

length (Figure 5.10). The phylogeny of HRV-A and HRV-C sequences in the putative recombinant region (3268-3525) was remarkably incongruent with that observed in flanking regions of the same length. All 10 HRV-C sequences grouped within the HRV-A clade in a similar manner similar to the 5ˈUTR region. However, in the putative recombinant 2A region, all HRV-C sequences (including HRV-Cc strain GQ223228) were actually embedded within the HRV-A clade. HRV-B sequences retained a phylogenetic position distinct from both HRV-A and HRV-C throughout 2A.

Subsequently, 38 additional 2A sequences from HRV-C strains in Edinburgh were combined with 58 HRV-C complete polyprotein sequences (>90% complete across the region 3127 – 3552 numbered by EF582385) downloaded in 2012 to give a complete dataset consisting of 96 HRV-C 2A sequences, representative of 39 HRV-C types. Analysis additionally included 194 HRV-A complete polyprotein sequences for comparison. This showed similar features to the original analysis with only 10 HRV-C complete genomes. All HRV-C sequences from all HRV-C types grouped in several distinct clades within the HRV-A clade within the putative recombinant region (3127 – 3552), implying the occurrence of several separate recombination events, similar to that observed in the 5ˈUTR. However, bootstrap values for inter-type relationships were generally poor for trees constructed from both datasets and so recombination parents could not be determined. Despite poor bootstrap support for inter-type relationships within the region, the grouping of HRV-C within HRV-A is well supported in both datasets (bootstrap values of 99 and 80 respectively).

FIGURE 5.10: Neighbour joining phylogenetic trees showing the putative recombinant portion (3267 – 3525) of the 2A coding region as compared with 5' (3006 – 3266) and 3' (3526 – 3786) flanking regions. *Numbering is from sequence EF582385 (HRV-C4). HRV-C sequences are labelled in black and HRV-A are labelled in white. The accession numbers of HRV-A/-B variants have been prefixed with their corresponding serotype designation. Branches are scaled by genetic distance.*

Of the 96 HRV-C strains with available sequences for the 2A region, 73 also had a 5'UTR sequence that was complete to position 179 (including all 38 strains isolated sequenced in this study). We had previously reported a monophyletic group of HRV-C sequences that was maintained in both the 5'UTR and 2A region (including types HRV-C1, -C3, -C6, -C7 and –C11) (McIntyre et al., 2010) and speculated that recombination in the 5'UTR and 2A regions may be linked. In order to investigate this potential linkage, we examined phylogenetic trees constructed from the same sequence set in both regions (Figure 5.11). This showed that, contrary to the pattern observed in 5'UTR region, HRV-Cc sequences did not group together in 2A and were interspersed throughout the HRV-C tree (marked in red on Figure 5.11). In the larger dataset, two clades containing 6 and 8 HRV-C types respectively were notably contiguous between the 5'UTR and 2A regions (marked with dotted lines and grey arrows in Figure 5.11). However, inspection of phylogenetic trees constructed from the VP1 region of HRV-C revealed that these groupings were additionally maintained in the coding region. In addition, one further bootstrap supported clade in 2A contained both HRV-Ca and HRV-Cc sequences and was present in the VP1 region. Indeed, when HRV-C sequences are considered alone, no phylogeny violations were recorded between VP1 and the putative recombinant 2A region (data not shown). This suggests that the recombination events observed in the 5'UTR and 2A regions are not, in fact, linked.

As there were no non-recombinant HRV-C 2A sequences, it was not possible to determine putative recombination breakpoints via GroupScan. However, inspection of the amino acid alignment of HRV-A, -B and –C sequences in this region revealed an area of obvious similarity corresponding closely with the region identified by divergence scanning (Figure 5.12). Annotation of the sequence alignment with known features of the 2A protein revealed the putative breakpoint to fall almost exactly at the boundary of the C-terminal domain (Petersen et al., 1999).

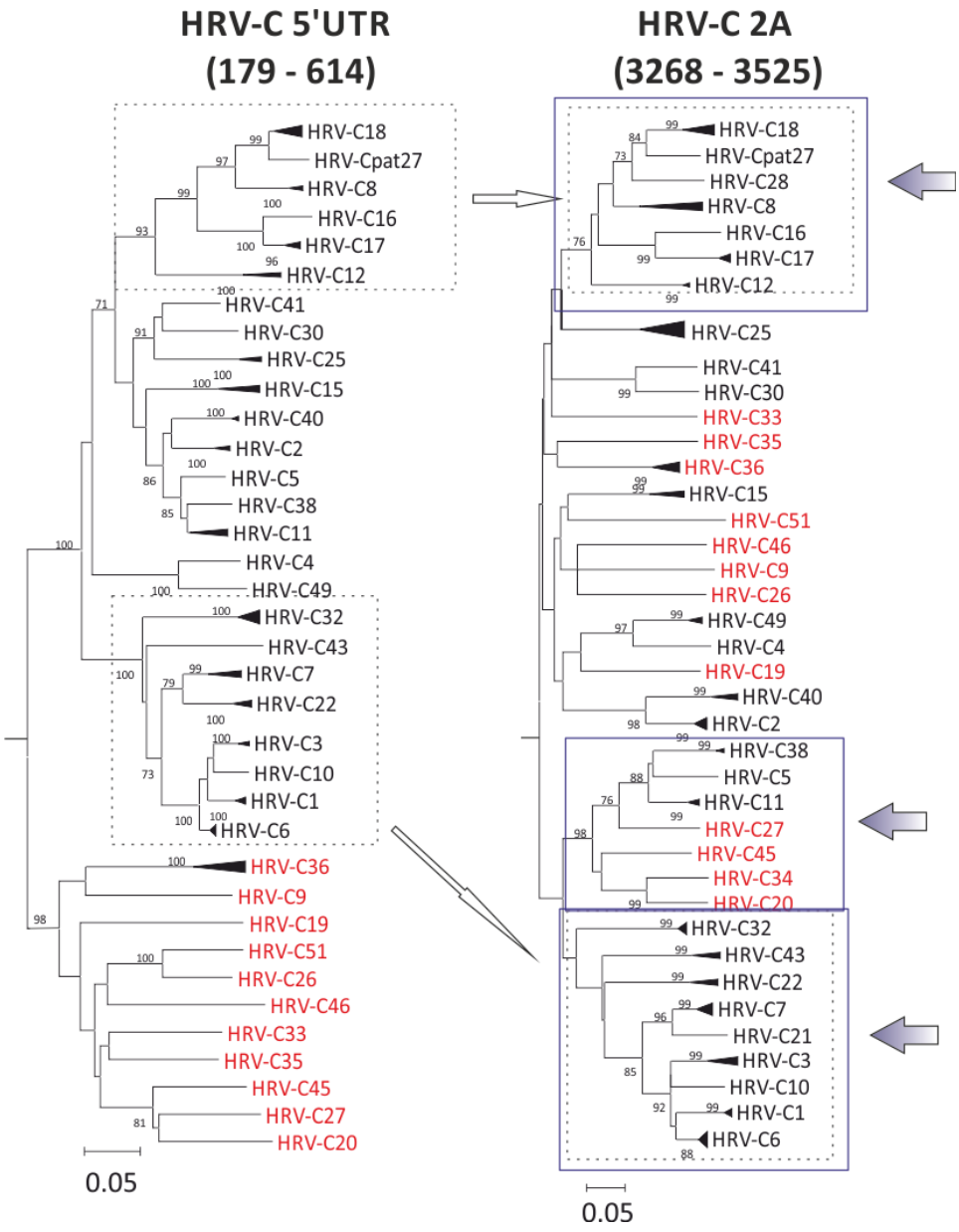


FIGURE 5.11: Neighbour joining phylogenetic trees showing the 5'UTR and recombinant 2A regions of HRV-C sequences. *Numbering is as shown and sequences are numbered by EF582385. Sequences which are non-recombinant in the 5'UTR region (HRV-Cc) are labelled in red type. Bootstrap supported clades containing more than two HRV-C types which are maintained in 5'UTR and recombinant 2A region are marked with white arrows and dotted lines. Bootstrap supported clades containing more than two HRV-C types which are also present in the VP1 region are marked with blue arrows and blue boxes. Branches are scaled by genetic distance.*

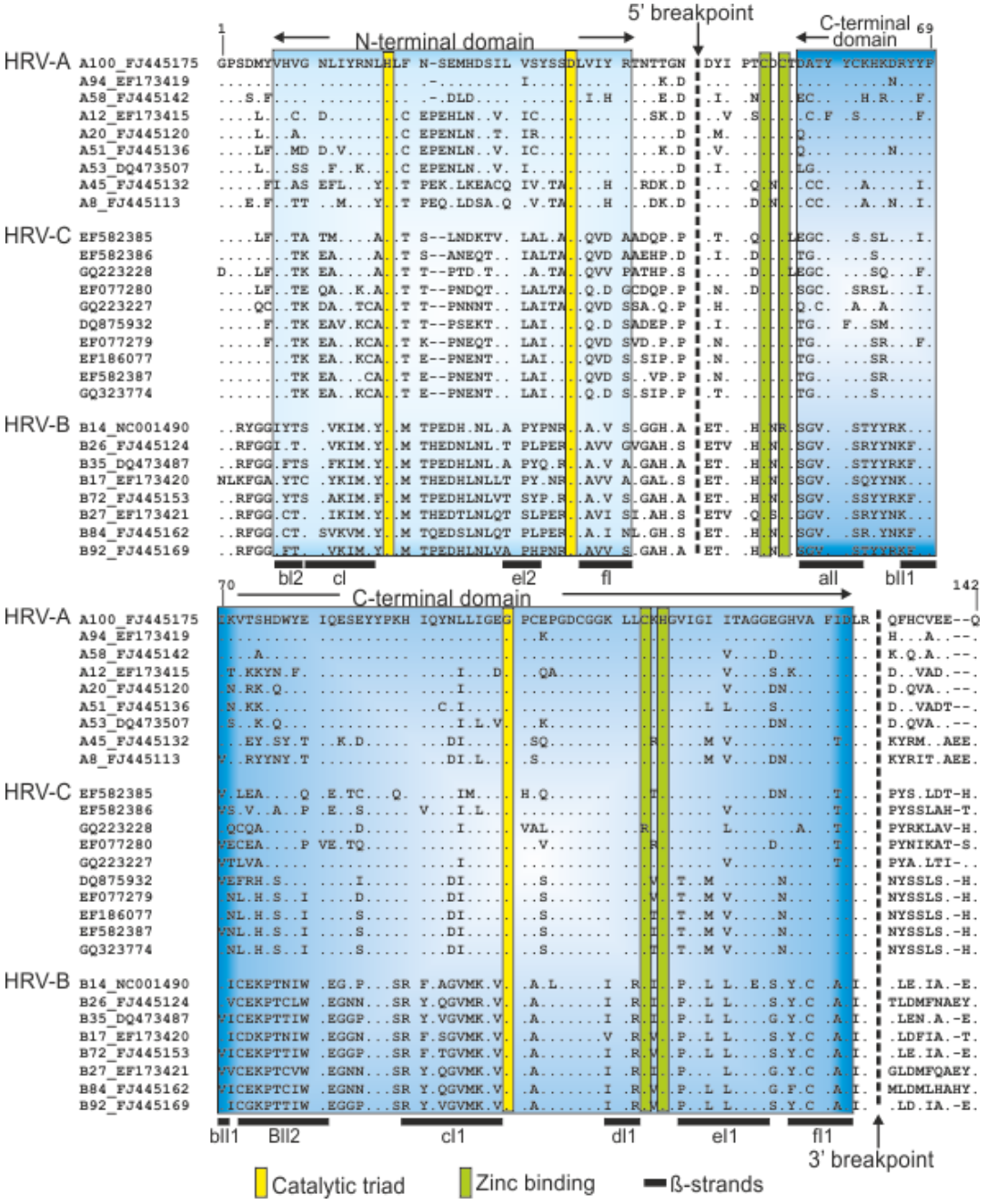


FIGURE 5.12: Sequence alignment of the 2A region of selected HRV-A, -B and –C sequences. *HRV-A (n=9) and HRV-B (n=8) sequences were selected on the basis of highest observed genetic divergence among prototype strains. HRV-C (n=10) strains include all of the originally published full genomes. Secondary structure elements are superimposed and annotated* (Petersen et al., 1999). *Inferred positions of recombination are shown by dotted lines (positions 3268 – 3525). Sequences are numbered by HRV-A2 : X02316. B-strands and catalytic/zinc binding residues are as previously proposed* (Petersen et al., 1999).

The putative recombinant region contained all of the zinc binding sites known within this protein (Petersen et al., 1999). Additionally, the proposed recombination breakpoint at position 3268 corresponded closely with the 3' end of the proposed *cis-acting* replication element (cre) for HRV-A, which has been previously mapped to between positions 3226 – 3270 (Gerber et al., 2001).

5.4 Discussion

5.4.1 Absence of recombination within the coding region of HRV-C

In the study of the phylogenetic relationships between different coding regions of the HRV-C genome, the most striking finding was the lack of evidence for recombination within the coding region. This was manifest as almost complete phylogenetic congruence between two capsid regions (VP4/VP2 and VP1) and a sequence fragment from the 3' end of the genome (3Dpol) (Figure 5.1). This was in stark contrast to the frequent recombination observed in all three species of the closely related human EV (Lukashev, 2005; Oprisan et al., 2002; Simmonds and Welch, 2006) and indeed in many other *Picornavirus* genera (Simmonds, 2006; Heath et al., 2006; Agol, 1997; Calvert et al., 2010). However, recombination has not been frequently observed within the coding region of other HRV species (Kistler et al., 2007b; Lewis-Rogers et al., 2009; Palmenberg et al., 2009).

One putative recombination event was noted within the coding region of HRV-C isolates. Sequence GQ223227 (N4) changed tree position between VP1 and 3Dpol, while the other member of the same type (Resp_10300) maintained a congruent position relative to other sequences (Figure 5.2). While we cannot definitively exclude the possibility that recombination occurred within a natural site located in this region and gave rise to a viable HRV-C strain, it seems relatively unlikely that this putative recombination event is genuine. Subsequent analyses on the occurrence of recombination in all three species of HRV indicated that the sporadic occurrence

of recombination within HRV appeared to predominantly involve full HRV types, rather than individual sequences (McIntyre et al., 2013b). It is possible that this apparently chimaeric genome has arisen due to contamination in laboratory analysis or sequence assembly. We would suggest that re-sequencing this isolate may be warranted to verify its ancestry. One subsequent study of recombination in HRV species showed a grossly elevated recombination frequency in HRV-C due to the inclusion of this aberrant sequence (Linsuwanon et al., 2011). In light of this, we have excluded it from this and further analyses of recombination in HRV-C sequences. With this sequence excluded, we found no other evidence of recombination occurring between HRV-C types within the coding region.

In studies of circulating EV strains, recombination frequency is correlated with both geographical and temporal separation of analysed strains (McWilliam Leitch et al., 2009a). In fact, even within these highly recombinant species, analysed sequences of both EV-A and EV-B were only observed to be 100% recombinant when 10 years had elapsed (Simmonds and Welch, 2006). The sequences produced in this study were all isolated from a single geographical location and covered a 2-3 year time period (from 2006 – 2008). With the inclusion of the ten available published full genome sequences, this time period was increased only to eight years. This raises the possibility that the detection of recombination may have been limited by the included dataset, especially if recombination occurs comparatively more rarely than in EV. In one study of EV strains isolated from Edinburgh blood donors in 2000-2001, no recombination among the study strains of EV-A was documented (Simmonds and Welch, 2006). However, when EV-A prototype strains and published strains covering a much larger time span were included, a high frequency of phylogenetic incongruity and putative recombination was observed. Further studies undertaken within this thesis investigated the occurrence of recombination within the coding region of HRV strains of all three species spanning 14 years and two geographically separate locations (Chapter 6).

5.4.2 Recombination within the 5ˈUTR of HRV-C

In accordance with previous studies (Huang et al., 2009; Wisdom et al., 2009a), we noted considerable phylogenetic incongruity between the 5ˈUTR of HRV-C and the rest of the genome downstream (VP4/VP2, VP1 and 3Dpol fragments). The majority of HRV-C sequences grouped within the HRV-A clade in the 5ˈUTR region (Figures 5.3 and 5.4). The 5ˈUTR/P1 junction is a known recombination hotspot in other members of the genus *Enterovirus* (Santti et al., 1999).

All HRV-C sequences included in this study maintained their phylogenetic grouping within their assigned type in the 5ˈUTR. Previously published analyses showed two HRV-C types which contained strains with both recombinant (HRV-Ca) and non-recombinant (HRV-Cc) 5ˈUTR sequences (Wisdom et al., 2009a) (marked as Ca/Cc on Table 5.2). One of these groups contained only sequences isolated from Edinburgh; three of which were HRV-Cc (Resp_2514, Resp_3938 and Resp_3102) and only one (Resp_3856) was HRV-Ca. Unfortunately, the aberrant strain Resp_3856 was missing and could not be further investigated. In the second Ca/Cc group, all published sequences from a Chinese study were reportedly HRV-Ca (Huang et al., 2009) and all those from Edinburgh samples were HRV-Cc. We did not find evidence of any other HRV-C types with both HRV-A and HRV-C like 5ˈUTR sequences. However, the previously reported existence of these Ca/Cc groups may indicate that contemporary recombination can occur within the 5ˈUTR region of HRV-C.

We observed the majority of recombination breakpoints within the 5ˈUTR to fall into two hotspot regions; near position 565 in the PPT and 523 in stem-loop 5 of the IRES. This is a considerably more restricted distribution of breakpoints than was previously reported (Huang et al., 2009). However, we were able to utilise a different scanning method (GroupScan) and a much larger dataset. Similarly to the coding region of HRV-C, the existence of a biological compatibility barrier and

requirement to preserve essential RNA secondary structure may generally limit the generation of viable recombinants in the 5'UTR to sequence surrounding these hotspot regions. The 5'UTR of HRV contains the highly structured and well conserved IRES and 5' cloverleaf, which have essential functions in viral replication (Palmenberg et al., 2009, 2010) and are also conserved in EV (Pöyry et al., 1996). Recombination breakpoints occurring at the hotspot regions described were found to have no influence on predicted RNA secondary structure (McIntyre et al., 2010). The observation that recombination is more likely to occur in areas of concentrated RNA secondary structure may be potentially supported by our observations (Runckel et al., 2013). However, this assertion remains controversial in EV (Simmonds and Welch, 2006) and the location of putative recombination breakpoints in HRV-C may be simply secondary to the highly conserved sequence in these regions. Putative recombination events observed in the 5'UTR of EV-C109 resulted in preservation of both the 5' cloverleaf and IRES (Yozwiak et al., 2010). The conservation of RNA secondary structure within recombinant 5'UTR sequences suggests a degree of functional constraint on the biological viability of chimaeric genomes within this region.

We also found no evidence of the putative second recombination event that was proposed at the 5' end of the 5'UTR (Huang et al., 2009), instead noting the swap of the entire 5'UTR in a modular fashion. Our methods for detection of the 5' breakpoint were potentially limited by the use of a 200 base fragment size in GroupScan analysis. However, the additional inspection of phylogenetic trees constructed from both 50 and 100 base regions at the 5' end of the UTR showed no evidence for grouping of any HRV-Ca sequence out with the HRV-A clade.

In EV-A71, recombination in the 5'UTR has been linked to epidemic spread. Specifically, a single epidemic genotype in the Netherlands has been found to have same recombinant 5'UTR sequence as an outbreak in Asian-Pacific region caused by

a different genotype (van der Sanden et al., 2011). HRV-A/-C 5ˈUTR recombinants are likely to have originally had increased viral fitness, as there are multiple VP4/VP2 lineages in the recombinant 5ˈUTR clade which indicates the occurrence of multiple recombination events (Figure 5.3). In addition, there is not a single recorded instance of an HRV-A variant possessing a HRV-C like 5ˈUTR sequence. However, no difference is observed between HRV-Ca and –Cc strains in terms of epidemiology or clinical presentation (Wisdom et al., 2009a). Schibler and co-workers found that artificially generated chimaeric genomes with a coding region derived from HRV-A16 and 5ˈUTR regions from other *Enterovirus* species (specifically HRV-Ca, HRV-Cc, HRV-B and EV-A) were all able to replicate successfully *in vitro* (Schibler et al., 2012). However, each recombinant genome was consistently out-competed by the wild type HRV-A16 in competition experiments. This indicates that, in this case, while the 5ˈUTR and coding regions of these viruses are functionally compatible, the wild type maintains a replication advantage.

5.4.3 Putative recombination events within the 2A coding region of HRV-C

The 2A coding region in HRV/EV encodes a chymotrypsin-like proteinase ($2A^{pro}$), of 142 amino acids and the crystal structure has been solved (Petersen et al., 1999). $2A^{pro}$ has two primary functions within the HRV life cycle. Firstly, it is responsible for the initial cleavage of the viral genome post-translation. $2A^{pro}$ cleaves the structural and non-structural proteins, separating the two domains at its own amino terminus (N-terminus) (Toyoda et al., 1986). In addition, $2A^{pro}$ is responsible for the cleavage of eukaryotic initiation factor eIF4G which is required for the recruitment of capped mRNA to the ribosome and therefore essential for host-cell mRNA translation (Glaser et al., 2003; Lloyd et al., 1988). This disruption of host cell metabolism ("host cell shut-off") allows uninhibited initiation of translation of picornavirus mRNA through an internally initiated process (via the IRES).

Structurally, $2A^{pro}$ is comprised of N-terminal (amino) and C-terminal (carboxyl) domains, connected by an approximately 12 amino acid interdomain loop (annotated onto sequence data in Figure 5.12) (Petersen et al., 1999). Unexpectedly, we observed a possible second site of HRV inter-species recombination manifest as a 260 base fragment of high sequence similarity between HRV-A and HRV-C. A similar observation was previously made in analysis of sequence GQ223227 (N4), which appeared to be more closely related to HRV-A sequences in the 3' end of 2A (Huang et al., 2009). Analysis of our expanded dataset (containing 96 HRV-C 2A sequences) revealed that the putative recombinant region corresponded precisely to the boundaries of the C-terminal domain. Phylogenetic trees constructed of the putative recombinant region (nucleotide positions 3268 – 3525) showed all known HRV-C sequences grouping within the HRV-A clade. The C-terminal domain of the $2A^{pro}$ is highly conserved, not only among HRV, but among all picornaviruses (Petersen et al., 1999). The observation of a putative recombination breakpoint in this region raises the possibility of a degree of modularity between the two domains. The N-terminal region is much more variable between HRV-A and HRV-C and the two species show segregation on phylogenetic trees, as is observed in other coding regions (Figure 5.10). Interestingly, a previous study observed this same modularity between the two domains when comparing proteinase sequences of HRV and *Streptomyces griseus* proteinase B (SGPB) by structure based alignment. Although the N-terminus was variable and significantly truncated in HRV, the C-terminal domain of both distantly related proteins was topologically identical (Petersen et al., 1999). This suggests a strong evolutionary constraint related to the structure of this type of proteinase, which may account for the observed sequence similarity among closely related species.

Inspection of the amino acid alignment for the 2A region of HRV-A, -B and –C highlighted an area of obvious similarity corresponding closely to the C-terminal domain of the 2A protein (Figure 5.12). HRV-B sequences remained divergent from

both HRV-A and HRV-C throughout the 2A region, consistent with their separate phylogenetic grouping (Figure 5.10). Analysis of pairwise nucleotide p-distances within the 2A coding region showed that there was considerably less genetic diversity observed within the putative recombinant region. This was true of intra-species distances for HRV-A and HRV-C, as well as inter-species distances between HRV-A and HRV-C. This observation was even more marked when pairwise amino acid p-distances were considered (Figure 5.13). In fact, the observed amino acid p-distance between HRV-A and HRV-C sequences in the non-recombinant region is more than twice that observed in the putative recombinant region.

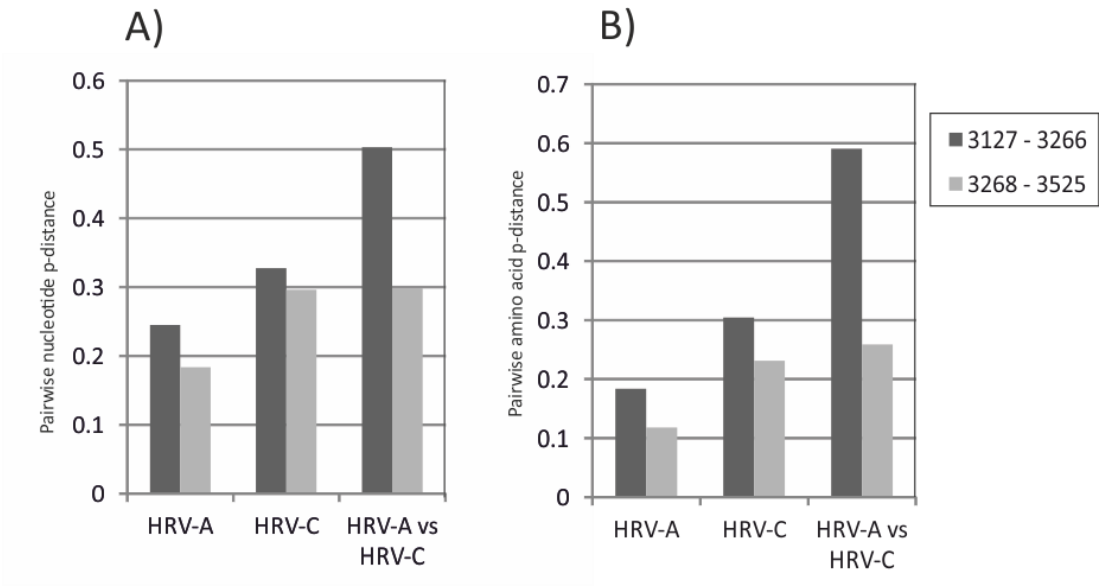


FIGURE 5.13: Pairwise nucleotide and amino acid p-distances for the 2A coding region of HRV-A, HRV-C and HRV-A/-C. *The 2A region has been divided into the non-recombinant (3127-3266) and putative recombinant (3267-3525) regions. Pairwise nucleotide p-distance is shown in A and pairwise amino acid p-distance is shown in B.*

The C-terminal domain of $2A^{pro}$ is structurally composed of 6 β-sheets which co-ordinate the binding of a zinc ion adjacent to the catalytic site of the protease. The zinc ion is extremely tightly bound within the protein, requiring denaturation of the protease in order to chelate it (Sommergruber et al., 1994). Despite a lack of direct involvement in enzyme catalysis, the presence of zinc within the molecule has been

found to be essential for function of the enzyme and indeed for viral replication (Glaser et al., 2003), secondary to a role in maintaining structural integrity (Voss et al., 1995). All four of the known zinc binding sites are located within the putative recombinant region of $2A^{pro}$ and consist of three cysteine (Cys) residues and one histidine (marked on Figure 5.12). The zinc binding sites are conserved in a broad range picornaviruses (Petersen et al., 1999) and in all sequenced HRV strains, with only one exception (GQ223228). This single sequence contains an Arginine (Arg) residue at position 3456, in place of Cys, corresponding to a single nucleotide mutation within the coding sequence. Given the differing physicochemical properties of these two amino acids, it seems unlikely that this substitution would allow efficient zinc binding. Additional strains of HRV-C8 that were sequenced in this region displayed a perfectly conserved Cys residue at the position in question. This single base change observed in only one published strain may be due to ambiguities often observed with certain sequencing techniques, where misincorporation of a nucleotide can lead to infrequent random base mutations.

Recombination is detected by the observation of the loss of genetic diversity and incongruence of phylogenetic relationships within a putative recombinant region. We did observe a higher than expected degree of sequence conservation within the C-terminal domain of HRV-A and HRV-C sequences. Our original analysis of putative recombination within this region suggests that HRV-C sequences were scattered throughout the HRV-A clade and that the conservation of certain monophyletic groups of sequences may indicate linkage between the recombination observed in the 5'UTR and 2A regions (McIntyre et al., 2010). However, the inclusion of almost 10 times the number of sequences included in the original analysis revealed that all HRV-C sequences grouped together within the HRV-A clade and that bootstrap supported clades maintained between the 5'UTR and 2A regions were also maintained throughout the coding region. Additionally, we observed largely congruent phylogenetic relationships between sequences in the VP1

and 2A coding regions. This suggests that a strong purifying selection influence in this region, rather than inter-species recombination, may be driving the pattern of genetic diversity observed. In order to maintain phylogenetic grouping within the coding region, a recombination event in this region would have presumably been required to have taken place prior to the diversification of HRV-C into genetically distinct types. The markedly reduced amino acid divergence, as compared with nucleotide divergence in the C-terminal domain of this protein is consistent with a high number of synonymous mutations, again suggesting a region under the influence of strong negative selection.

Consistent with the proposed functional and structural modularity between these two domains is the fact that the 5' end of the putative recombinant region corresponds almost exactly to the 3' end of the HRV-A *cis-acting* replication element (cre) (Gerber et al., 2001). A breakpoint in this region would not disrupt the secondary structure or function of the cre, which is essential for RNA replication. Additionally, the fact that there is no structural homology between the proposed HRV-A cre and the corresponding sequence fragment in HRV-C is consistent with the current understanding that the HRV-C cre is located within the VP2 region (Cordey et al., 2008).

In conclusion, recombination within the coding region appears to be extremely rare within HRV-C. This observation is limited by the fact that our analysis of HRV-C sequences was restricted to strains that were mostly isolated from a small geographical region and over a restricted time period. Subsequent analysis focussed on the occurrence of recombination within all species of HRV over a significantly extended time period and isolated from several distinct geographical regions. The occurrence of recombination within the 5'UTR of HRV-C sequences has been previously documented (Huang et al., 2009; Wisdom et al., 2009b; Palmenberg et al., 2009). However, to our knowledge, this was the first study to attempt to accurately

pinpoint recombination breakpoints. The grouping of recombinant HRV-Ca sequences with HRV-A at the extreme 5' end of the 5'UTR suggests a modular exchange of almost the entire 5'UTR region. Sequence analysis of full HRV-C genomes revealed a highly conserved sequence fragment corresponding to the C-terminal domain of the 2A proteinase. However, we did not find evidence for a linkage of the recombination events between the 5'UTR and 2A regions.

Chapter 6
Recombination in the evolution of HRV genomes

6.1 Introduction

The occurrence of recombination within the coding region of all three species of HRV is generally considered to be both rare and sporadic, notwithstanding significant disagreement about its actual frequency within the currently published literature. For example, a study of 45 HRV-A and HRV-B serotypes found some evidence of phylogenetic incongruity within the coding region of the genome, but could not confirm any putative recombination events by RDP analysis (Lewis-Rogers et al., 2009). In contrast, a study of 35 full genomes of HRV-A/-B found that the full genome phylogeny was identical to that constructed from various subgenomic regions and the sequence identity throughout the genome was largely consistent (Kistler et al., 2007b). However, several small scattered putative recombination events were identified by RDP analysis. Phylogenetic incongruity was detected between VP4/VP2 and 3Dpol sequences of HRV-A/-B (Savolainen et al., 2004). However, due to the poor bootstrap support for inter-type relationships in the VP4/VP2 region, none of these potential incongruities could be further analysed. Sporadic putative recombination events have also been identified within the coding region of HRV genomes (Tapparel et al., 2009b). Finally, a large study which included at least one sequenced isolate of all 100 known HRV-A/-B serotypes found that the full genome phylogeny recovered did show a number of inconsistencies with subgenomic regions (Palmenberg et al., 2009) and ultimately suggested that several contemporary HRV types had arisen due to recombination. The limitation in previous studies may be related to the sample size or time span of data used for the analysis. None of the studies mentioned included both a representative set of full genome sequences and several sequenced examples from most HRV types.

Our analysis (presented in Chapter 5) of a set of contemporary HRV-C sequences was limited by the absence of sequences collected prior to 2001, giving only a 7 year time span for analysis. Analysis of several extensively recombinant EV-B types (E9, E11 and E30) indicated that even these required sequence analysis spanning several years for recombination to become widely evident (McWilliam Leitch et al., 2009a, 2010). In particular, E11 showed a half-life for individual recombinant forms of nearly 10 years. An analysis of HRV-C sequences over an extended time period had never been conducted and in addition, HRV-A/-B have not been extensively analysed using a large dataset of both contemporary and prototype sequences. Therefore, an extended HRV dataset was constructed in order to more definitively measure recombination frequencies occurring during the diversification and contemporary evolution of HRV types. The analysis presented within this chapter focussed on examination of the occurrence of recombination within all published full genome sequences of all three species of HRV and additionally compared sequenced isolates spanning 14 years. This study was the first to comprehensively characterise all detectable recombination events in HRV and was published in 2013 (McIntyre et al., 2013b) .

6.2 Materials and Methods

6.2.1 Sample selection

28 HRV positive respiratory samples isolated from 1995 – 1997 in Finland were provided by Carita Savolainen-Kopra (National Institute for Health and Welfare (THL), Mannerheimintie 166, Finland). The VP4/VP2 sequences for these isolates had been previously published and were within the series of accession numbers EU590043 – EU590113 (Savolainen-Kopra et al., 2009a) and AY015114 – AY015174 (Savolainen et al., 2002). By analysis of pairwise nucleotide p-distances and phylogenetic clustering, these isolates were divided into 22 HRV types (10 HRV-A, 6 –B and 6 –C).

41 HRV positive respiratory samples isolated in Edinburgh of these same 22 HRV types were identified and selected. In addition, a further 23 HRV-C sequences relating to 11 previously provisionally assigned HRV-C types were identified during the course of this work and were also included in the recombination analysis.

6.2.2 Amplification of three coding regions of HRV

The VP4/VP2 and VP1 regions of HRV-A, -B and –C were amplified as previously described (Chapter 2). A fragment of 3Dpol stretching from positions 6414 – 6896 in HRV-A (numbered by FJ445111- HRV-A1), 6475 – 6960 in HRV-B (numbered by X01087 : HRV-B14) and 6361 – 6835 in HRV-C (numbered by EF582385 : HRV-C4) was amplified as described in Chapter 2 using primers listed (Appendix 1)[18]. The region described above corresponded to that amplified in HRV-C in previous work (McIntyre et al., 2010). Where necessary, for samples which proved challenging to amplify, a SuperScript III (Invitrogen, UK) protocol was used as previously described.

Two HRV-B isolates could only be amplified by using a multiplexed mixture of primers for all three HRV species. Subsequent inclusion of inner primers for all species in the sequencing reaction gave a clear, readable sequence. Neither of these two HRV-B sequences were found to be recombinant in the 3Dpol region.

All sequences included in this study have been submitted to Genbank and assigned accession numbers in the series KC342054 – KC342173.

6.2.3 Construction of datasets for recombination analysis

Two separate datasets were constructed for analysis of recombination within HRV genomes. The first consisted of all VP4/VP2, VP1 and 3Dpol sequences generated in this study (McIntyre et al., 2013b), in our previous work on recombination within

[18] Within this chapter, all analysis of the 3Dpol region was undertaken using the fragment between the co-ordinates described above.

HRV-C (McIntyre et al., 2010) and extracted from all HRV polyprotein sequences and population sets available on Genbank (downloaded on 6/6/12). In order to meet the criteria for inclusion in the study, sequences were required to be more than 90% complete across the VP4/VP2 region previously described (615 – 1004 numbered by EF582385), the full VP1 region and the 3Dpol fragment described above. In addition, all sequences which had been annotated as non-functional or contained stop codons were excluded. This resulted in inclusion of 136 HRV-A, 53 HRV-B and 121 HRV-C sequences, in which the three regions were concatenated to form a single long sequence suitable for recombination analysis. The second dataset constructed included all available HRV polyprotein sequences that were more than 90% complete across the P1, P2 and P3 regions (numbered by Genbank annotation for FJ445111 : HRV-A1, X01087 : HRV-B14 and EF582385 : HRV-C4).

6.2.4 Likelihood mapping analysis

Many inferred phylogenetic clades observed in preliminary studies suffered from poor bootstrap support and this was especially marked when considering deep phylogenetic relationships within the VP4/VP2 and 3Dpol regions. In order to assess whether conclusions based on inspection of bootstrap supported phylogenetic clades were valid, it was necessary to confirm the presence of adequate phylogenetic signal and that a tree-like topology could adequately describe the phylogenetic relationships within the dataset.

Likelihood mapping analysis was undertaken using TREE-PUZZLE v5.2 (Schmidt et al., 2002), using 10000 random quartets, uniform rates and a specified HKY substitution model. All other parameters were estimated from the data by maximum likelihood during execution of the program. Fragment datasets were analysed for each studied region and each species separately. The analysis was also carried out in duplicate to ensure convergence of results.

6.2.5 Analysis of mean substitution rates

A total of 18 HRV-A and 23 HRV-C sequences from 6 distinct HRV types were selected for analysis of mean substitution rates. These included sequences representing HRV-A9, -A28, -A78, -C9, -C12 and –C18. These were selected on the basis of a lack of observed inter- and intra-type recombination between VP1 and 3Dpol. Datasets were constructed of each individual type group and subjected to analysis with SBP (Kosakovsky Pond et al., 2006a), GARD (Kosakovsky Pond et al., 2006b) and RDP (Martin et al., 2010) prior to inclusion.

Analysis of mean substitution rates was carried out in BEAST (Drummond and Rambaut, 2007). All analyses were carried out using the SRD06 model of nucleotide substitution, as recommended for protein coding sequences (Shapiro et al., 2006). A relaxed uncorrelated lognormal molecular clock and a constant population size were utilised and all other priors were optimized during execution of the program. The MCMC chain was run for 20 million states and the output was recorded every 1000 states. Trees were directly compared to those produced by bootstrap re-sampled MCL neighbour joining methods to ensure consistency of the phylogenetic relationships inferred.

6.2.6 Recombination analysis

In addition to phylogenetic analysis carried out by visual inspection of bootstrap supported clade members, branch length and overall tree topology, several computational methods were employed to both screen entire datasets for evidence of recombination and also to further characterise any putative recombination events.

Datasets containing both full HRV species and single HRV types were subjected individually to analysis with GARD (Kosakovsky Pond et al., 2006b), SBP (Kosakovsky Pond et al., 2006a) and the RDP v4.0 software package (Martin et al., 2010). For any putative recombination event highlighted on large-scale screening or

inspection of phylogenetic trees, additional confirmatory analysis was undertaken with RDP v4.0 programs. In each instance, a putative recombinant sequence or group of sequences were screened along with the nearest neighbours in each region. As a control measure, several groups of sequences which were observed to have no evidence of phylogenetic incongruence on inspection of bootstrap supported phylogenetic trees were analysed by RDP, GARD and SBP. All analyses confirmed no evidence of recombination within these datasets.

Further characterisation of putative recombination events was undertaken using the GroupScan program within the SSE v1.0 software package (Simmonds, 2012) and supplemented by construction and inspection of bootstrapped phylogenetic trees of the regions flanking every putative recombinant region. Unlike the other computational methods employed, GroupScan specifically scores the extent of phylogenetic clustering of a query sequence with specified control groups and therefore takes changes in bootstrap supported phylogenetic tree topology into account. This is especially useful in cases where parental sequences are not present in the dataset and phylogenetic inference is more likely to determine an accurate grouping than sequence distance alone. Additionally, the usage of full clades as control groups prevents the unnecessary aggregation of valuable sequence data that occurs when creating a consensus sequence. As GroupScan outputs a series of grouping score values calculated over a specified sliding window and plotted as a linear graph, the putative recombination breakpoint can be easily calculated by the intersection of the lines. In these studies, concatenated sequence fragments from VP4/VP2, VP1 and 3Dpol were scanned with a fragment size of 300 bases and an increment of 30. In each case, the query sequence was the potential recombinant and it was scanned against the full nearest neighbour group in both VP1 and 3Dpol. Each analysis was additionally repeated with a fragment size of 600 bases and a window of 30 to ensure consistency. This analysis gave highly similar results.

6.3 Results

6.3.1 Phylogenetic analysis within HRV type groups

Sequenced HRV isolates were combined with sequences obtained from Genbank to give a total of 164 HRV-A, 70 –B and 145 –C sequences that were more than 90% complete across the VP4/VP2, VP1 and 3Dpol regions described. In addition, a second dataset was assembled containing all published HRV polyprotein sequences that were more than 90% complete across the P1, P2 and P3 regions (201 HRV-A, 71 –B and 59 –C).

Inspection and comparison of phylogenetic trees constructed for the VP4/VP2 and VP1 regions revealed the majority of HRV sequences grouping in both regions into bootstrap supported phylogenetic clusters, analogous to type groups[19]. The putative type groupings defined by VP1 clustering and pairwise nucleotide p-distance analysis were largely maintained in both the VP4/VP2 and full P1 regions.

Bootstrap support observed for clusters of HRV types within all three species was generally poor within the short and well conserved VP4/VP2 region (Figure 6.1). Only two HRV types showed evidence for any phylogenetic incongruity. HRV-A31 and HRV-A47 have previously been described as distinct HRV-A types, due to their genetic distance and phylogenetic relationship in VP1 (Chapter 4). However, on inspection of VP4/VP2 trees, contemporary sequences group separately and are not easily assigned to either type. This anomaly has been previously described as a pair of HRV types which are less divergent in VP4/VP2 than would be expected. Likewise, HRV-A29 and HRV-A44 are equally related to one contemporary sequence JN815252 in the VP4/VP2 region. This serotype pair was previously highlighted as potentially able to be combined.

[19] Although we have proposed that certain HRV type groups undergo revisions to their classical definitions (for example, HRV-A29 and HRV-A44), all analysis presented within this chapter follows the conventional labelling of HRV types.

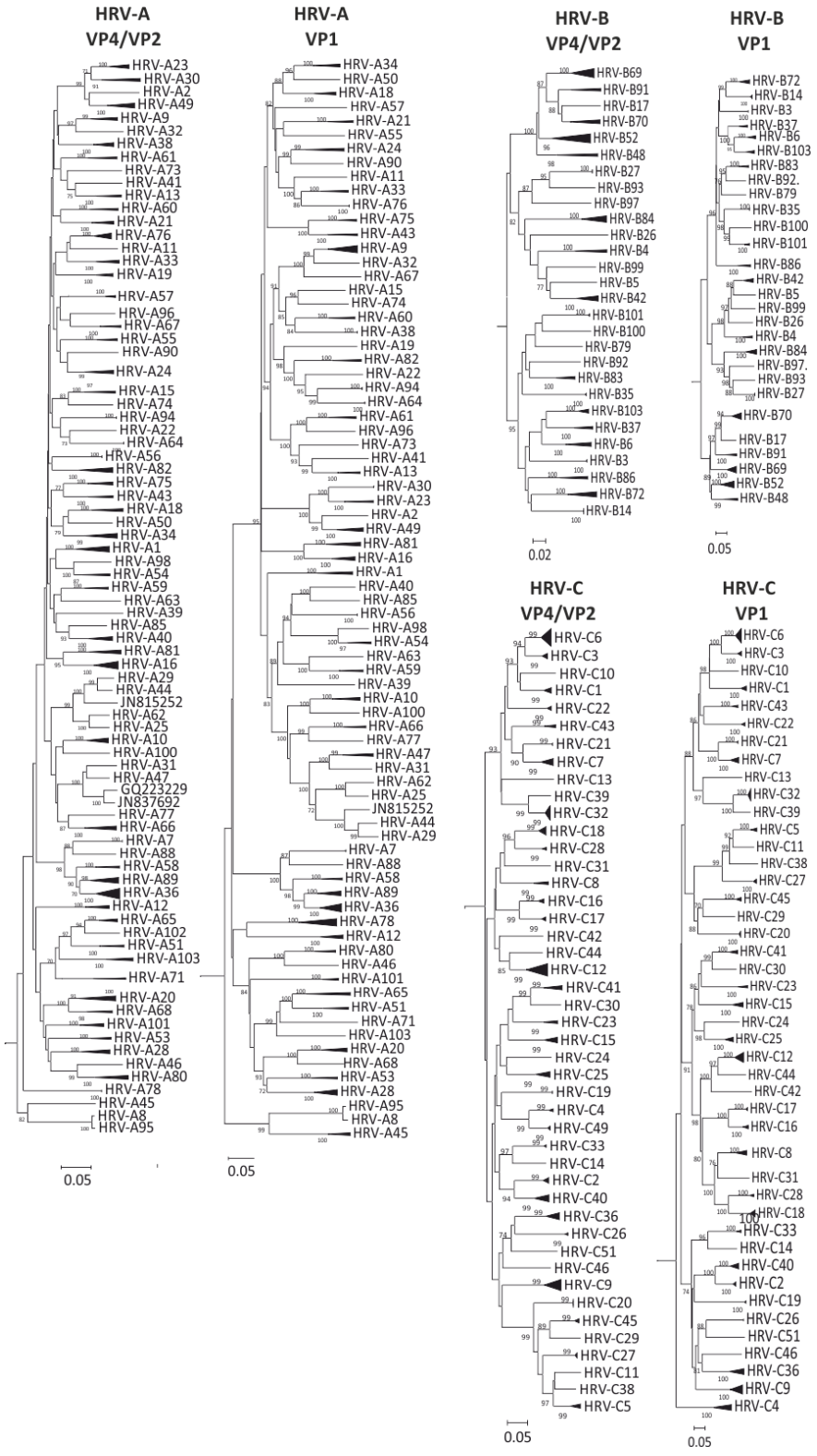

HRV-A
VP4/VP2
HRV-A23
HRV-A30
HRV-A2
HRV-A49
HRV-A9
HRV-A32
HRV-A38
HRV-A61
HRV-A73
HRV-A41
HRV-A13
HRV-A60
HRV-A21
HRV-A76
HRV-A11
HRV-A33
HRV-A19
HRV-A57
HRV-A96
HRV-A67
HRV-A55
HRV-A90
HRV-A24
HRV-A15
HRV-A74
HRV-A94
HRV-A22
HRV-A64
HRV-A56
HRV-A82
HRV-A75
HRV-A43
HRV-A18
HRV-A50
HRV-A34
HRV-A1
HRV-A98
HRV-A54
HRV-A59
HRV-A63
HRV-A39
HRV-A85
HRV-A40
HRV-A81
HRV-A16
HRV-A29
HRV-A44
JN815252
HRV-A62
HRV-A25
HRV-A10
HRV-A100
HRV-A31
HRV-A47
GQ223229
JN837692
HRV-A77
HRV-A66
HRV-A7
HRV-A88
HRV-A58
HRV-A89
HRV-A36
HRV-A12
HRV-A65
HRV-A102
HRV-A51
HRV-A103
HRV-A71
HRV-A20
HRV-A68
HRV-A101
HRV-A53
HRV-A28
HRV-A46
HRV-A80
HRV-A78
HRV-A45
HRV-A8
HRV-A95
0.05
HRV-A
VP1
HRV-A34
HRV-A50
HRV-A18
HRV-A57
HRV-A21
HRV-A55
HRV-A24
HRV-A90
HRV-A11
HRV-A33
HRV-A76
HRV-A75
HRV-A43
HRV-A9
HRV-A32
HRV-A67
HRV-A15
HRV-A74
HRV-A60
HRV-A38
HRV-A19
HRV-A82
HRV-A22
HRV-A94
HRV-A64
HRV-A61
HRV-A96
HRV-A73
HRV-A41
HRV-A13
HRV-A30
HRV-A23
HRV-A2
HRV-A49
HRV-A81
HRV-A16
HRV-A1
HRV-A40
HRV-A85
HRV-A56
HRV-A98
HRV-A54
HRV-A63
HRV-A59
HRV-A39
HRV-A10
HRV-A100
HRV-A66
HRV-A77
HRV-A47
HRV-A31
HRV-A62
HRV-A25
JN815252
HRV-A44
HRV-A29
HRV-A7
HRV-A88
HRV-A58
HRV-A89
HRV-A36
HRV-A78
HRV-A12
HRV-A80
HRV-A46
HRV-A101
HRV-A65
HRV-A51
HRV-A71
HRV-A103
HRV-A20
HRV-A68
HRV-A53
HRV-A28
HRV-A95
HRV-A8
HRV-A45
0.05
HRV-B
VP4/VP2
HRV-B69
HRV-B91
HRV-B17
HRV-B70
HRV-B52
HRV-B48
HRV-B27
HRV-B93
HRV-B97
HRV-B84
HRV-B26
HRV-B4
HRV-B99
HRV-B5
HRV-B42
HRV-B101
HRV-B100
HRV-B79
HRV-B92
HRV-B883
HRV-B35
HRV-B103
HRV-B37
HRV-B6
HRV-B3
HRV-B86
HRV-B72
HRV-B14
0.02
HRV-B
VP1
HRV-B72
HRV-B14
HRV-B3
HRV-B37
HRV-B6
HRV-B103
HRV-B83
HRV-B92.
HRV-B79
HRV-B35
HRV-B100
HRV-B101
HRV-B86
HRV-B42
HRV-B5
HRV-B99
HRV-B26
HRV-B4
HRV-B84
HRV-B97.
HRV-B93
HRV-B27
HRV-B70
HRV-B17
HRV-B91
HRV-B69
HRV-B52
HRV-B48
0.05
HRV-C
VP4/VP2
HRV-C6
HRV-C3
HRV-C10
HRV-C1
HRV-C22
HRV-C43
HRV-C21
HRV-C7
HRV-C13
HRV-C39
HRV-C32
HRV-C18
HRV-C28
HRV-C31
HRV-C8
HRV-C16
HRV-C17
HRV-C42
HRV-C44
HRV-C12
HRV-C41
HRV-C30
HRV-C23
HRV-C15
HRV-C24
HRV-C25
HRV-C19
HRV-C4
HRV-C49
HRV-C33
HRV-C14
HRV-C2
HRV-C40
HRV-C36
HRV-C26
HRV-C51
HRV-C46
HRV-C9
HRV-C20
HRV-C45
HRV-C29
HRV-C27
HRV-C11
HRV-C38
HRV-C5
0.05
HRV-C
VP1
HRV-C6
HRV-C3
HRV-C10
HRV-C1
HRV-C43
HRV-C22
HRV-C21
HRV-C7
HRV-C13
HRV-C32
HRV-C39
HRV-C5
HRV-C11
HRV-C38
HRV-C27
HRV-C45
HRV-C29
HRV-C20
HRV-C41
HRV-C30
HRV-C23
HRV-C15
HRV-C24
HRV-C25
HRV-C12
HRV-C44
HRV-C42
HRV-C17
HRV-C16
HRV-C8
HRV-C31
HRV-C28
HRV-C18
HRV-C33
HRV-C14
HRV-C40
HRV-C2
HRV-C19
HRV-C26
HRV-C51
HRV-C46
HRV-C36
HRV-C9
HRV-C4
0.05

FIGURE 6.1: Neighbour joining phylogenetic trees showing the VP4/VP2 and VP1 region of HRV-A, -B and –C. *Analysis included full VP1 region as previously described and VP4/VP2 region between 615 – 1004(numbered by EF582385). Branches are scaled by genetic distance.*

No evidence of recombination was evident within the capsid region on inspection of phylogenetic trees[20] (Figure 6.1). In addition, analysis with RDP, GARD and SBP failed to highlight any significant instances of putative recombination.

6.3.2 Likelihood mapping analysis

Likelihood mapping analysis revealed strong support for a tree-like pattern of evolution in all three gene regions of all three species (Figure 6.2). In TREE-PUZZLE v5.2 output, the proportion of analysed quartets with a reliably resolved tree-like phylogeny is mapped into the three vertices of the triangle, while the proportion of quartets which produce a star-like phylogeny or phylogenetic noise are mapped in the centre. For each species and each analysed gene, the proportion of unresolved quartets did not exceed 1%. In contrast, general support for a resolved bifurcating phylogenetic tree consistently exceeded 90%. It can therefore be inferred that conclusions based on analysis of bootstrap supported phylogenetic trees are likely to be valid.

[20] The sequence JX291115 which was reported as a potential recombinant between VP4/VP2 and VP1 in the analysis of HRV-C taxonomy was not included in this analysis as the sequence was published after the datasets for recombination analysis were assembled.

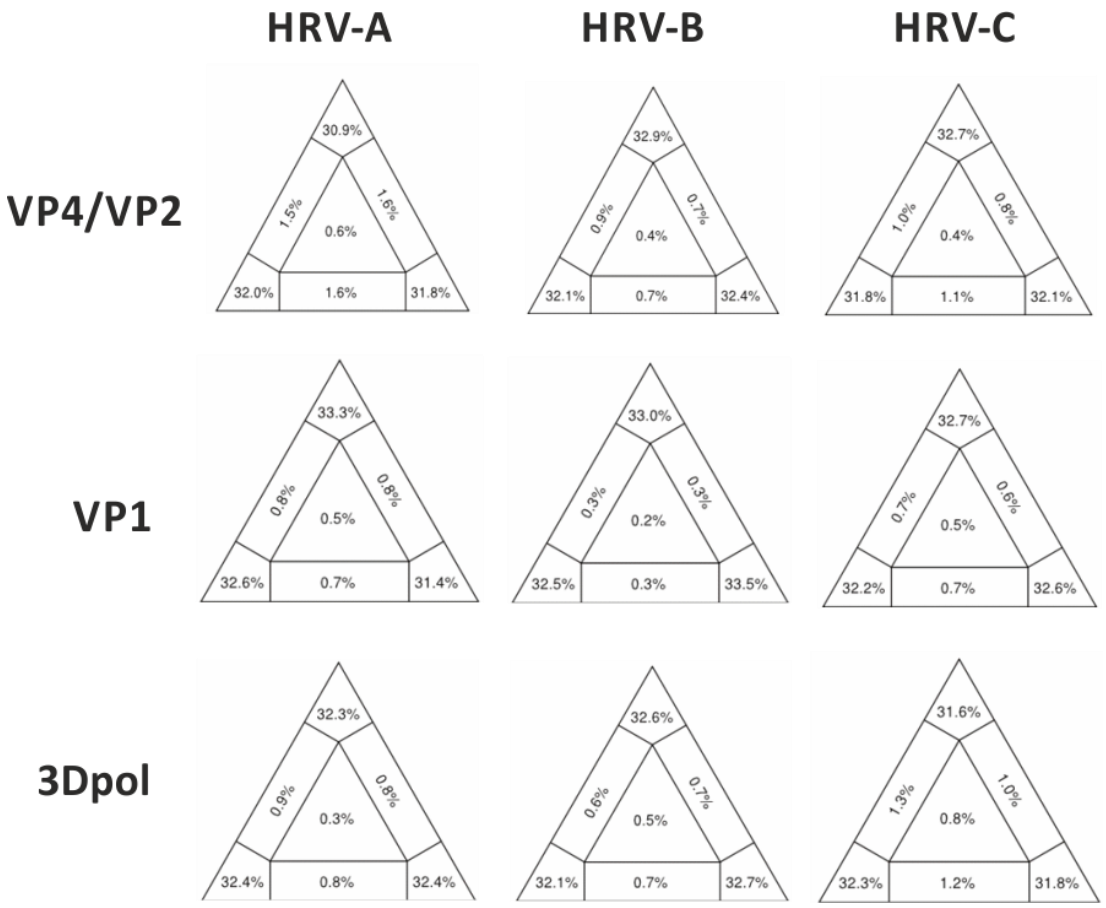


FIGURE 6.2: Likelihood mapping analysis of the VP4/VP2, VP1 and 3Dpol regions of HRV-A, -B and –C. *Proportion of quartets in the vertices represent those with fully resolved bifurcating phylogeny, whereas the proportion in the centre represent those with star-like or network-like evolution.*

6.3.3 Analysis of mean substitution rates

Mean substitution rates were calculated for the VP1 and 3Dpol regions of three HRV-A and HRV-C types. Phylogenetic trees analysed for each type showed no bootstrap supported change in topology within each type or species. In addition, datasets including sequences from either a single type or a single HRV species showed no evidence of recombination when analysed with RDP, GARD and SBP. The total number of HRV-A and HRV-C sequences analysed was 18 and 23 respectively and the datasets spanned 40 years (HRV-A) and 14 years (HRV-C).

Insufficient sequences of individual HRV-B types were available for inclusion in the analysis.

Calculated mean substitution rates and 95% highest posterior density (HPD) intervals were of the order of 10^{-3} to 10^{-2} substitutions per site per year. The rates observed were largely similar between both regions and both HRV species analysed (Figure 6.3). Although both the VP1 and 3Dpol regions of HRV-C showed a 2 to 3-fold elevation in substitution rates (6.9138 x 10^{-3} and 9.8491 x 10^{-3} respectively) when compared to those estimates obtained for HRV-A strains (3.9954 x 10^{-3} and 3.4586 x 10^{-3}), the 95% HPD intervals (which are analogous to confidence intervals) were clearly overlapping between both species.

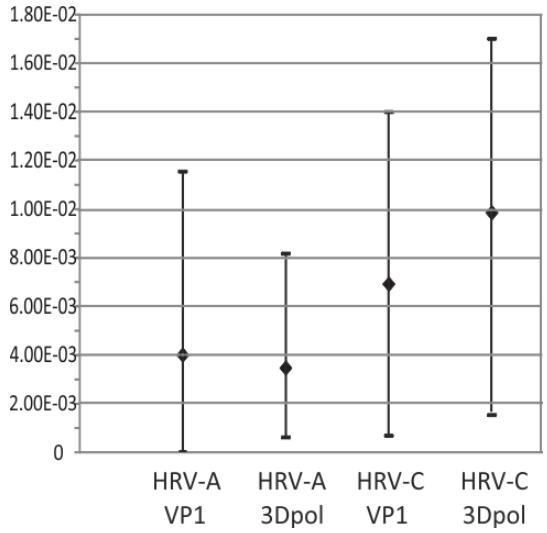


FIGURE 6.3: Mean substitution rates and 95% highest posterior density intervals for two coding regions of HRV-A and HRV-C.

Phylogenetic trees produced by Bayesian methods were identical in branching order and type membership to the bootstrap resampled MCL neighbour joining trees utilised in subsequent analysis (data not shown). Both independent executions of the BEAST program gave highly similar results. In addition, analyses run with

sequence data excluded or dates of isolation scrambled showed no overlap with the genuine estimates obtained.

6.3.4 Comparisons of pairwise nucleotide p-distances

The concurrence of similar mean substitution rates between the VP1 and 3Dpol regions of HRV-A and –C should create a linear relationship between sequence divergences between these two regions. This was indeed observed in both intra- and inter-type distance ranges for HRV-B and HRV-C (Figure 6.4). Within distributions for both HRV-B and HRV-C, there were very few, if any outlying data points representing potentially recombinant sequences and the line of best fit displayed a gradient of close to 1 (0.9042 and 0.9418 respectively). This correlation provides evidence for maintained equal substitution rates and a general lack of recombination throughout the period of diversification into these distinct types.

In contrast to the congruent relationship observed in HRV-B and –C sequences, HRV-A sequences were almost uniformly less divergent than expected in the 3Dpol region when compared to VP1 divergence. A line of best fit for the entire HRV-A dataset gave a gradient of 0.8714. In addition, a large number of outlying data points were observed which may be indicative of individual recombinant sequences during the diversification of HRV-A sequences.

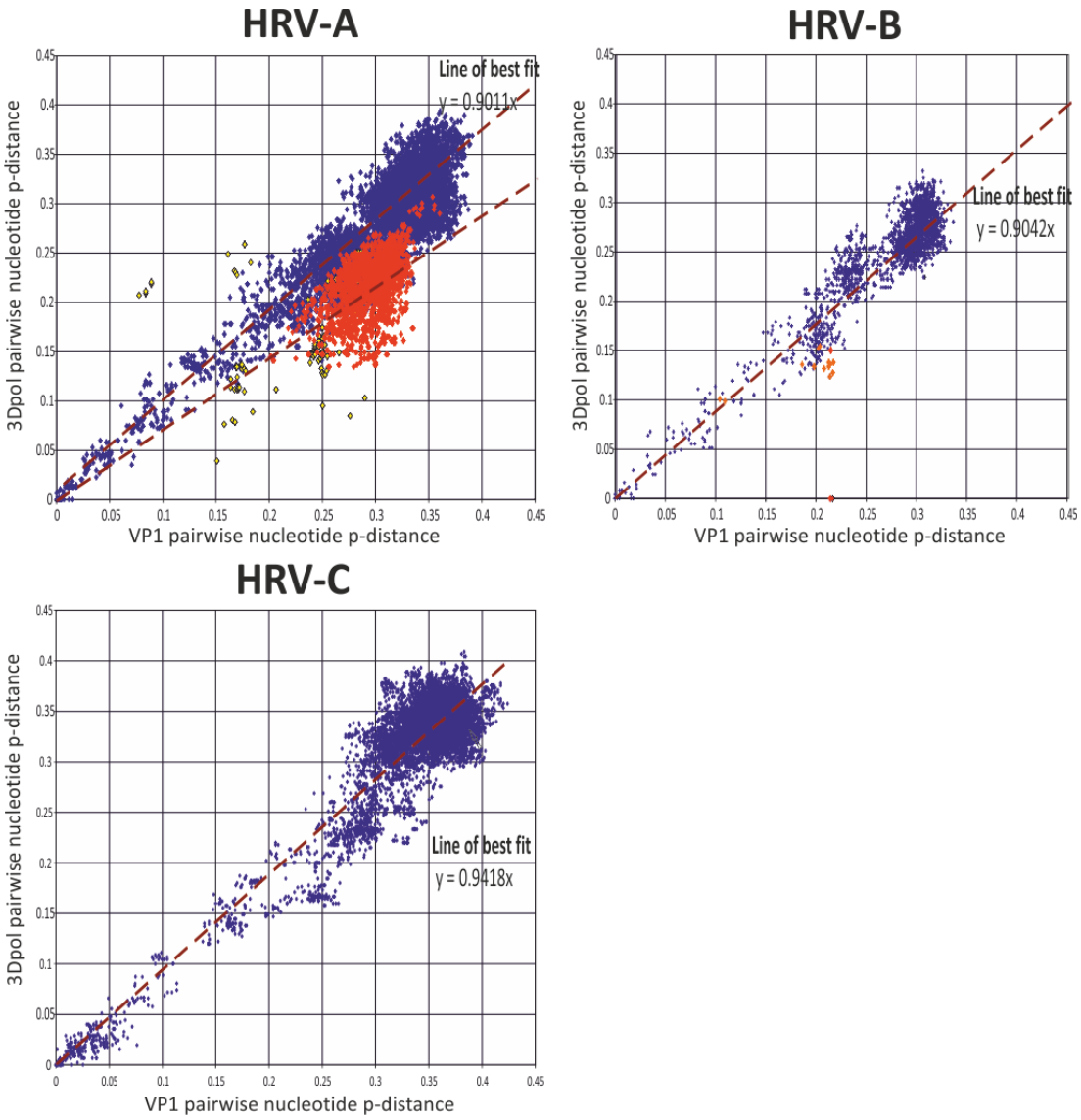


FIGURE 6.4: Comparison of pairwise nucleotide p-distances between the VP1 and 3Dpol regions of HRV-A, -B and –C. *In each distribution, the line of best fit is indicated by the red dotted line. Line is shown for putative recombinant and non-recombinant HRV-A sequences separately. The graph depicting HRV-A comparisons shows putative recombination events involving full HRV-A type groups in red and those involving single sequences in yellow. For HRV-B sequences, all putative recombination events are shown in red. The previously described, likely erroneous recombinant HRV-C sequence GQ223227, has been excluded from the analysis.*

Analysis of individual HRV-A pairwise p-distances that were discordant between genome regions allowed the identification of potential recombinant sequences. In combination with inspection of bootstrap supported phylogenetic trees, pairwise p-

distances of HRV-A sequences could be readily divided into three distinct sets. The majority of sequences were non-recombinant and showed no evidence of bootstrap supported changes in tree topology (shown in blue on Figure 6.4). Non-recombinant sequences had a line of best fit gradient of 0.9011 similar to that of species B and C. The remaining putative recombinant HRV-A sequences could be categorised into two distinct distributions; one arising from large groups containing several types (shown in red on Figure 6.4) and one containing instances of sporadic phylogenetic incongruity involving single sequences alone (shown in yellow on Figure 6.4). This analysis highlighted the possibility of significant recombination acting in both the recent diversification of HRV-A (sporadic recombinants) and more ancient events that occurred before the diversification of HRV-A into types.

6.3.5 Analysis of recombination within the coding region of HRV-A

With the exception of HRV-A12 and HRV-A78, all HRV-A sequences grouped into four bootstrap supported clades (numbered on Figure 6.5), which branched basally on both VP1 and 3Dpol trees. These basal clades were maintained in the VP1, 3Dpol, P1, P2 and P3 regions. Due to poor bootstrap support for relationships between distantly related HRV types, the basal clades were not maintained in VP4/VP2. However, no bootstrap supported incongruity was observed. The clades were additionally observed to be present, bootstrap supported and monophyletic in every other individual gene region of HRV-A (data not shown), with the exception of the short and well-conserved VP4 and 3B regions. Interestingly, although multiple instances of bootstrap supported recombination events were recorded within clades 1, 2 and 3, there were no instances of recombination within the coding region occurring between these clades. In addition, no changes in tree topology were observed with clade 4 due to the inclusion of only 3 HRV-A types within it.

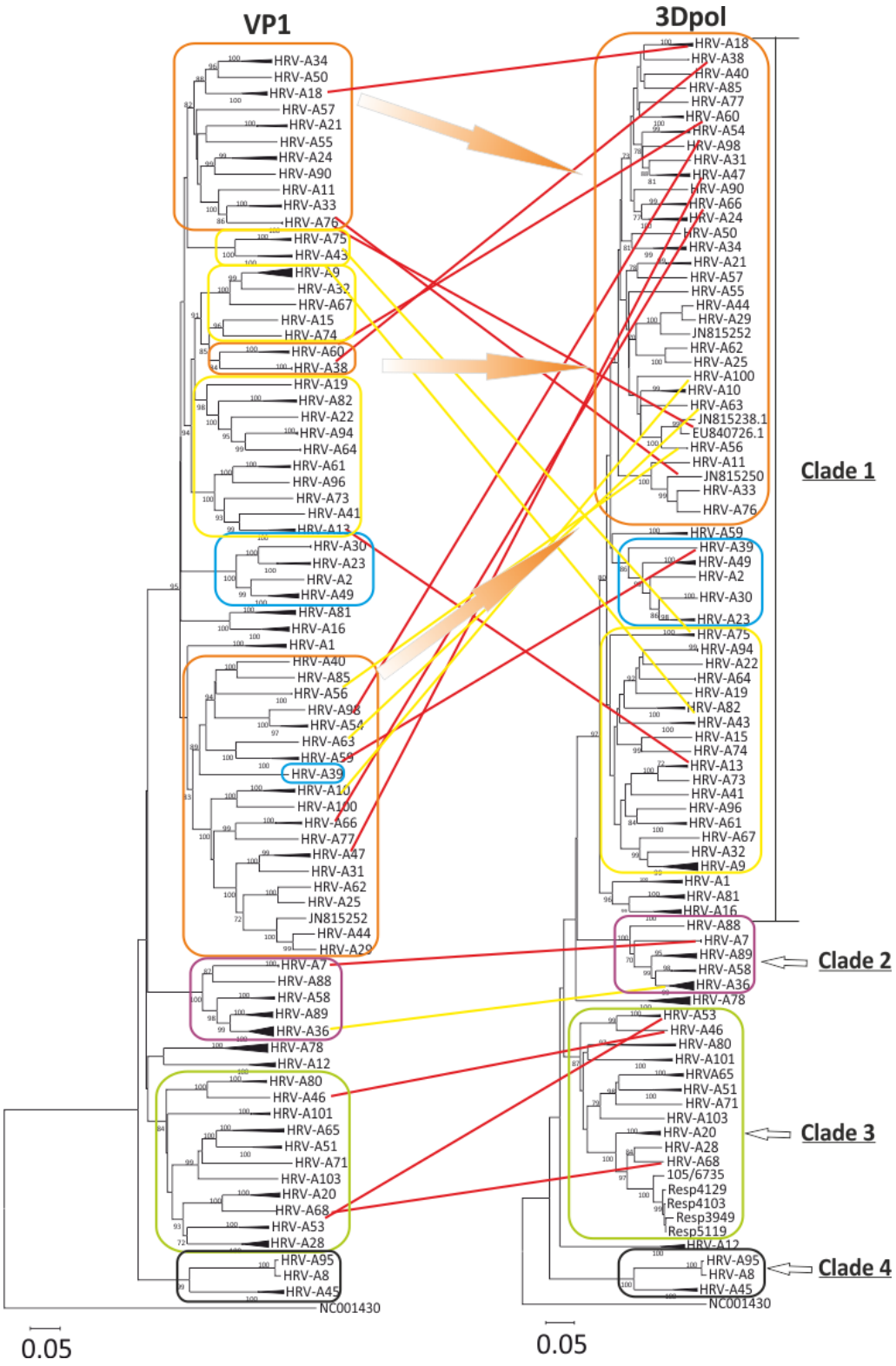

VP1
3Dpol
HRV-A34
HRV-A50
HRV-A18
HRV-A57
HRV-A21
HRV-A55
HRV-A24
HRV-A90
HRV-A11
HRV-A33
HRV-A76
HRV-A75
HRV-A43
HRV-A9
HRV-A32
HRV-A67
HRV-A15
HRV-A74
HRV-A60
HRV-A38
HRV-A19
HRV-A82
HRV-A22
HRV-A94
HRV-A64
HRV-A61
HRV-A96
HRV-A73
HRV-A41
HRV-A13
HRV-A30
HRV-A23
HRV-A2
HRV-A49
HRV-A81
HRV-A16
HRV-A1
HRV-A40
HRV-A85
HRV-A56
HRV-A98
HRV-A54
HRV-A63
HRV-A59
HRV-A39
HRV-A10
HRV-A100
HRV-A66
HRV-A77
HRV-A47
HRV-A31
HRV-A62
HRV-A25
JN815252
HRV-A44
HRV-A29
HRV-A7
HRV-A88
HRV-A58
HRV-A89
HRV-A36
HRV-A78
HRV-A12
HRV-A80
HRV-A46
HRV-A101
HRV-A65
HRV-A51
HRV-A71
HRV-A103
HRV-A20
HRV-A68
HRV-A53
HRV-A28
HRV-A95
HRV-A8
HRV-A45
NC001430
0.05
HRV-A18
HRV-A38
HRV-A40
HRV-A85
HRV-A77
HRV-A60
HRV-A54
HRV-A98
HRV-A31
HRV-A47
HRV-A90
HRV-A66
HRV-A24
HRV-A50
HRV-A34
HRV-A21
HRV-A57
HRV-A55
HRV-A44
HRV-A29
JN815252
HRV-A62
HRV-A25
HRV-A100
HRV-A10
HRV-A63
JN815238.1
EU840726.1
HRV-A56
HRV-A11
JN815250
HRV-A33
HRV-A76
Clade 1
HRV-A59
HRV-A39
HRV-A49
HRV-A2
HRV-A30
HRV-A23
HRV-A75
HRV-A94
HRV-A22
HRV-A64
HRV-A19
HRV-A82
HRV-A43
HRV-A15
HRV-A74
HRV-A13
HRV-A73
HRV-A41
HRV-A96
HRV-A61
HRV-A67
HRV-A32
HRV-A9
HRV-A1
HRV-A81
HRV-A16
HRV-A88
HRV-A7
HRV-A89
HRV-A58
HRV-A36
HRV-A78
Clade 2
HRV-A53
HRV-A46
HRV-A80
HRV-A101
HRVA65
HRV-A51
HRV-A71
HRV-A103
HRV-A20
HRV-A28
HRV-A68
105/6735
Resp4129
Resp4103
Resp3949
Resp5119
Clade 3
HRV-A12
HRV-A95
HRV-A8
HRV-A45
Clade 4
NC001430
0.05

FIGURE 6.5: Neighbour joining phylogenetic trees showing the VP1 and partial 3Dpol regions of all available HRV-A strains. *Where HRV-A types are maintained between both regions, these have been collapsed for ease of reference. Contiguous basal clades 1 to 4 are marked by coloured boxes. Clades 2, 3 and 4 are marked by purple, green and black boxes respectively, while Clade 1 is divided into sub-clusters which group together in 3Dpol. Within Clade 1, the orange arrows indicate the formation of one 3Dpol clade from three VP1 clades. Putative recombinant HRV-A types which undergo a bootstrap supported change in nearest neighbour between VP1 and 3Dpol are marked with a red line. Any changes in tree topology which are bootstrap supported only on P1, P2 and P3 trees are indicated with a yellow line. Branches are scaled by genetic distance.*

The majority of observed putative recombination events within HRV-A involved full groups of sequences, analogous to one HRV type undergoing a change in tree position between VP1 and 3Dpol. Within clades 2 and 3, four HRV-A types underwent a change in bootstrap supported nearest neighbour between VP1 and 3Dpol and an additional HRV-A type showed evidence of a putative recombination event that was evident on comparison of P1, P2 and P3 trees.

Clade 1 formed several distinct sub-clusters of HRV-A types and displayed a clear breakdown in the phylogenetic relationships between these types between VP1 and 3Dpol. One monophyletic, bootstrap supported clade in 3Dpol (marked in orange on Figure 6.5) was formed from members of three, distinct and distantly related VP1 clades. Relationships between types were largely incongruent within the newly formed 3Dpol clade and the comparisons of pairwise p-distances between sequences within this clade corresponded to the large cluster of putative recombinant sequences between 0.25 and 0.32 VP1 p-distances (coloured red in Figure 6.5). However, there was uniformly observed poor bootstrap support for nearest neighbour relationships between some HRV-A types. These poorly supported relationships were largely mirrored on P3 trees. Remarkably, the majority of observed phylogenetic incongruity within all HRV-A sequences corresponded to full HRV-A types and more contemporary relationships between strains of the same HRV-A type are perfectly congruent. This may suggest the occurrence of widespread recombination during the period of the diversification of HRV-A into the presently observed types. The number of bootstrap supported nodes that each putative recombinant group

violated in the transition between VP1 and 3Dpol tree position ranged from two to eight (summarised in Table 6.1).

TABLE 6.1: Phylogenetic incongruities observed between VP1 and 3Dpol in HRV-A sequences

HRV Type	**Accession Numbers**	**Nearest neighbour VP1**	**Nearest Neighbour 3Dpol**	**Bootstrap supported nodes violated[a]**	**Calculated breakpoint mean (range)**
A7	FJ445176 DQ473503	A88	A36 A58 A89[b]	3	6576
A13	FJ445116 FJ445117	A41	A73	2	3393 (3365 – 3420)
A18	FJ445118 ***F292/8643***[c] JF781496 **JF781510**[d]	A34 A50	A24 A31 A38 A40 A47 A54 A60 A66 A77 A85 A90 A98	7[e]	5276 (5024 -5577)
A31	FJ445126	A25 A29 A44 A47 A62 JN815252[f]	A47 A54 A98	6	5224
A38	FJ445180 DQ473495 **JQ994496** **JN541272**	A15 A60 A74	A18 A24 A31 A40 A47 A54 A60 A66 A77 A85 A90 A98	7[g]	5130 (5118 – 5143)

HRV Type	Accession Numbers	Nearest neighbour VP1	Nearest Neighbour 3Dpol	Bootstrap supported nodes violated[a]	Calculated breakpoint mean (range)
A39	AY751783	A40 A54 A56 A59 A63 A85 A98	A2 A23 A30 A49	4	3431
A47	FJ445133 JN837692 GQ223229	A25 A29 A31 A44 A62 JN815252	A31 A54 A98	6	5424 (5416 – 5435)
A46	DQ473506	A80	A53	4	3280
A53	DQ473507 JN798587	A28	A46	4	3213 (3198 – 3228)
A54	FJ445138	A40 A56 A85 A98	A31 A47 A98	6	3228
A60	FJ445143 JN798590	A15 A38 A74	A18 A24 A31 A38 A40 A47 A54 A66 A77 A85 A90 A98	8[h]	5117 (5111 – 5123)
A66	FJ445148 JN112340 **JN621246** **JQ837715**	A77	A24	6	5652 (5645 – 5674)
A68	FJ445150 **JN798578**	A20	A28	4	3998 (3995 – 4001)
A76 (except FJ445182, DQ473502)	EU840726 JN815328 **JX074055** **JX074049**	A76	A56	7	5130 (5129 – 5131)
A98	FJ445139 FJ445173	A40 A54 A56 A85	A31 A47 A54	6	3360 (3343 – 3377)

[a] This relates to the number of bootstrap supported nodes a group of sequences violates to form the new grouping.
[b] Where more than one type is listed as a single nearest neighbour, this represents the most closely related phylogenetic clade which is bootstrap supported. For example, in cases where the putative recombinant type is an outgroup to a clade or its relationship with the most closely related sequences is not supported, all bootstrap supported types were considered as nearest neighbour.
[c] Sequences indicated in bold italics are those which consist only of non-consecutive sequence fragments and are therefore not represented on P1, P2 and P3 trees. These were excluded from breakpoint analysis.
[d] Sequences indicated in bold font are those which are members of a type which is a putative recombinant but are not themselves present on VP1 and 3Dpol trees due to sequence gaps. These are however represented on P1, P2 and P3 trees.
[e] Due to a lack of bootstrap support for nearest neighbour grouping in 3Dpol for A18, nodes changes are taken from grouping with A34 to A85 (nearest bootstrap supported neighbour in the P3 region).
[f] JN815252 is part of a group of previously described contemporary sequences which cannot be definitively assigned to either the HRV-A29 or HRV-A44 classical groupings. It is therefore listed as a single sequence.
[g] Node changes for HRV-A38 between VP1 and 3Dpol are taken from grouping with HRV-A60 to HRV-A18.
[h] Node changes for HRV-A60 between VP1 and 3Dpol are taken from grouping with HRV-A38 to HRV-A77.

Several HRV-A types showed a pattern of divergence whereby contemporary strains grouped separately from older prototype strains in the 3Dpol region. This represents tentative evidence of recombination within the time-frame of this analysis. As previously reported (Tapparel et al., 2009b), contemporary HRV-A76 strains collected between 1999 – 2010 group separately from the HRV-A76 prototype strain and form a close association with the HRV-A56 group. HRV-A56 maintains a congruent phylogenetic position between VP1 and 3Dpol and so this suggests that HRV-A76 have potentially recombined with HRV-A56 between 1999 and 2010 (Figure 6.6 A1). The intra-type p-distance range of contemporary HRV-A76 strains compared with prototype strains is 0.084 to 0.093 in VP1 and 0.209 to 0.229 in 3Dpol. Unfortunately, an insufficient number of contemporary HRV-A76 strains were available to perform Bayesian analysis to determine the date of most recent common ancestor of this group.

HRV-A68 and HRV-A33 also displayed a similar pattern of divergence, whereby the older prototype strains grouped preferentially with HRV-A28 and HRV-A76 type

strains respectively in 3Dpol (Figure 6.6 A2 and A3). However, both of these events relate to putative recombination between types which are closely related and so their significance is uncertain. All of the above listed phylogenetic incongruities were confirmed as putative recombination events by analysis with SBP, GARD and RDP (data not shown).

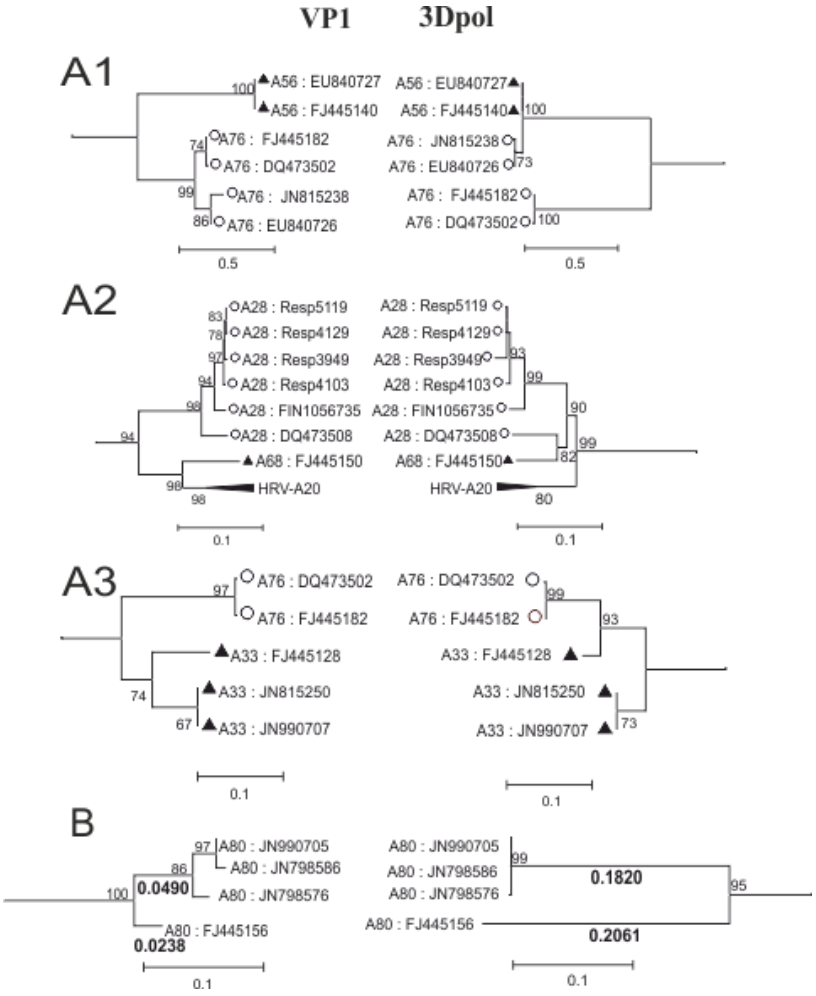


FIGURE 6.6: Neighbour joining phylogenetic trees showing distinct patterns of recombination observed in HRV-A sequences. *The branch to tree root has been collapsed for ease of reference and the outgroup used for analysis is given below. A: Phylogenetic trees showing examples of incongruent topology between VP1 and 3Dpol for (1): HRV-A56 and HRV-A76 (outgroup: HRV-A7), (2): HRV-A20, HRV-A28 and HRV-A68 (outgroup: HRV-A95) and (3): HRV-A33 and HRV-A76 (outgroup: HRV-A88). B: Phylogenetic tree showing branch length discrepancies between VP1 and 3Dpol within HRV-A80 (outgroup: HRV-A46). Branches are scaled by genetic distance.*

HRV-A80 sequences showed a substantial difference in intra-type branch lengths between the two regions. However, there was no accompanying change in tree topology (Figure 6.6 B). The HRV-A80 prototype strain (FJ445156) displayed a pairwise nucleotide p-distance from contemporary HRV-A80 strains of 0.078 in VP1 and 0.207 in 3Dpol. This discrepancy was confirmed with RDP analysis (data not shown).

6.3.6 Recombination between the P1, P2 and P3 regions of HRV-A

Sequence groupings observed in phylogenetic trees constructed for the whole P1, P2 and P3 coding regions were largely consistent with those observed in VP1 and 3Dpol. As expected, P1 was analogous to the VP1 tree and P2/P3 were similar to 3Dpol. The putative recombinant sequences groups observed in the 3Dpol tree were also evident upon inspection of P2 and P3 trees. However, seven further HRV-A types displayed evidence of bootstrap supported phylogenetic incongruity which was not bootstrap supported in 3Dpol (Figure 6.7). Of these seven further HRV-A types which were observed to undergo bootstrap supported changes in tree position, four showed evidence of incongruity between P1 and P2 and three between P2 and P3 (marked in yellow in Figure 6.7). All but one involved full HRV-A types (listed in Table 6.2).

The HRV-A40 type group includes three recently described polyprotein sequences. These three sequences were not included in the VP1 and 3Dpol analysis due to the presence of gaps which resulted in sequence completeness below the threshold of 90%. However, on analysis of the whole P3 region, the three contemporary HRV-A40 strains (JX074051, JQ245067 and JN798579) formed a bootstrap supported clade separate to the HRV-A40 prototype strain (marked in yellow in Figure 6.7). The prototype strain maintained its grouping with HRV-A85 (marked with black line in Figure 6.7).

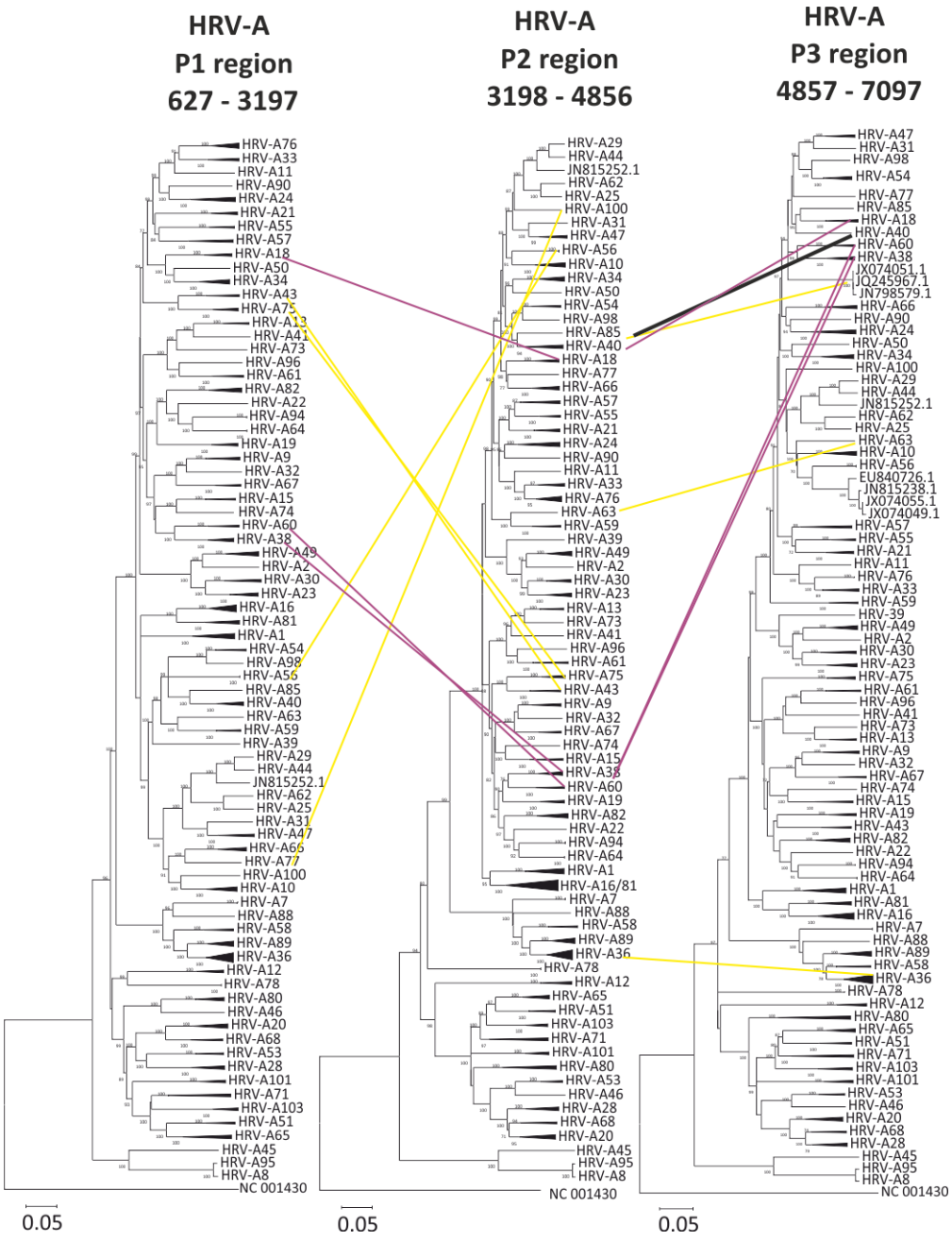


FIGURE 6.7: Neighbour joining phylogenetic trees showing the P1, P2 and P3 regions of HRV-A sequences. *P1, P2 and P3 regions are numbered by FJ445111 : HRV-A1. Seven HRV-A types which change bootstrap supported tree position between these regions and do not have a corresponding bootstrap supported phylogenetic incongruity on VP1 and 3Dpol trees are marked with yellow lines. The non-recombinant HRV-A40 prototype strain is marked with a black line. Those 3 HRV-A types which change bootstrap supported tree position both between P1/P2 and P2/P3 are marked with purple lines. Putative recombinant sequences which underwent one bootstrap supported change in tree topology between VP1 and 3Dpol are not shown. Branches are scaled by genetic distance.*

TABLE 6.2: Phylogenetic incongruities observed between the P1, P2 and P3 regions in HRV-A sequences which lack bootstrap support in the VP1 or 3Dpol regions

		Nearest Neighbours			**Number of changes**		
HRV Type	**Accession Numbers**	**P1**	**P2**	**P3**	**P1 → P2**	**P2 → P3**	**Calculated breakpoint mean (range)**
A36	DQ473505 ***F41598***[a] ***F44035*** JF781497 JN614994 **JN621243**[b] JN798583 JN798584 **JN815241** **JN815242** JN815246 JN837697 **JX074050**	A89	A89	A58	0	2	5068 (5049-5086)
A40 (except FJ445129)	**JN798579** **JQ245067** **JX074051**	A40	A40	A18 A31 A38 A40 A47 A54 A60 A77 A85 A98	1	0	N/A[c]
A43	DQ473510 ***F41576*** **JF781503** **JN837690**	A11 A18 A21 A24 A33 A34 A50 A55 A57 A75 A76 A90	A9 A15 A19 A22 A32 A64 A67 A74 A75 A82 A94	A9 A15 A19 A22 A32 A64 A67 A74 A75 A82 A94	3	0	3248 (3244-3257)

HRV Type	**Accession Numbers**	**P1**	**P2**	**P3**	**P1 → P2**	**P2 → P3**	**Calculated breakpoint mean (range)**
A56	FJ445140 EU840727	A40 A54 A85 A98	A10	A10 A76c[d]	8	0	3617
A63	FJ445146	A59	A59	A10 A56 A76c	0	6[e]	5124
A75	FJ445131 JN815237	A11 A18 A21 A24 A33 A34 A43 A50 A55 A57 A76 A90	A9 A15 A19 A22 A32 A43 A64 A67 A74 A82 A94	A9 A15 A19 A22 A32 A43 A64 A67 A74 A82 A94	3	0	3204 (3199 – 3209)
A100	FJ445175	A10	A25 A29 A44 A62 JN815252	A10 A25 A29 A44 A56 A62 A63 A76c JN815252	5	0	3292

[a] Sequences indicated in bold italics are those which consist only of non-consecutive sequence fragments and are therefore not represented on P1, P2 and P3 trees.
[b] Sequences indicated in bold are those which are not present on VP1 and 3Dpol trees due to sequence gaps.
[c] As the new tree position of the contemporary HRV-A40 sequences is as an outgroup to a clade which contains the original HRV-A40 prototype sequence and has poor bootstrap support for deep phylogenetic relationships, it was not possible to determine an accurate breakpoint for these sequences. However, by phylogenetic anlaysis, it can be assumed that this breakpoint lies between the P2 and P3 regions.
[d] A76c represents contemporary HRV-A76 sequences (EU840726, JN815238, JX074055 and JX074049).
[e] Node changes for A63 between P2 and P3 are taken from A59 to A10 and A56.

In addition, three HRV-A types were observed to undergo two separate changes in bootstrap supported nearest neighbour, with a change occurring between P1 and P2 and again between P2 and P3 (listed in Table 6.3). The closely related sequence pair

HRV-A38/A60 underwent all changes together, suggesting that these events may have occurred before their diversification into phylogenetically distinct types. The breakpoint analysis was achieved by considering both HRV types as a single query sequence group. All recombination events described were confirmed by analysis with RDP, SBP and GARD software programs (data not shown).

TABLE 6.3: HRV-A types which displayed evidence of phylogenetic incongruity between both the P1/P2 and P2/P3 regions

		Nearest neighbour			**Number of changes**		
HRV Type	**Accession Number**	**P1**	**P2**	**P3**	**P1 → P2**	**P2 → P3**	**Calculated breakpoint mean (range)[a]**
A18	FJ445118 ***F292/8643*** JF781496 **JF781510**	A34 A50	A66 A77	A85	9	5	3397 (3328 – 3526)
A38	DQ473495 FJ445180 **JN541272** **JQ994496**	A9 A15 A32 A60 A67 A74	A19 A60	A18 A31 A40 A47 A54 A60 A77 A85 A98	4 [b]	14[c]	3235 (3207 – 3285)
A60	FJ445143 JN798590	A9 A15 A32 A38 A67 A74	A19 A38	A18 A31 A38 A40 A47 A54 A77 A85 A98	3	10	3257 (3218 – 3296)

[a] The calculated mean breakpoint given corresponds to that one which was not observed on VP1 and 3Dpol trees and not described in Table 6.1.

[b] Node changes for HRV-A38/A60 between P1 and P2 are taken from HRV-A60/A38 to HRV-A19

[c] Node changes for HRV-A38/A60 between P2 and P3 are taken from HRV-A60/A38 to HRV-A40

Comparisons of pairwise nucleotide p-distances between the P1/P2 and P2/P3 regions of HRV-A mirrored the patterns observed upon inspection of phylogenetic

trees. In particular, the majority of the recombination events involving full HRV-A type groups were detected upon comparison of P1/P2 (coloured red in Figure 6.8), including the large scale re-organisation of inter-type relationships seen within Clade 1 (Figure 6.5). A line of best fit imposed on the P1/P2 distribution showed a gradient of 0.7341 for putative recombinant comparisons and 0.9799 for non-recombinant, similar to the pattern observed for the analysis conducted between VP1 and 3Dpol sequences. In contrast, the comparison of P2/P3 revealed relatively fewer, sporadic recombination events. These included instances where contemporary sequences of a particular type group separately to the type strain and are more distantly related than would be expected. For example, the relationship between contemporary HRV-A76 strains and the HRV-A56 type group and the putative recombination event observed within HRV-A80. A line of best fit for the P2/P3 distribution showed a gradient of 1.2082 for putative recombinant comparisons and 0.9813 for non-recombinant.

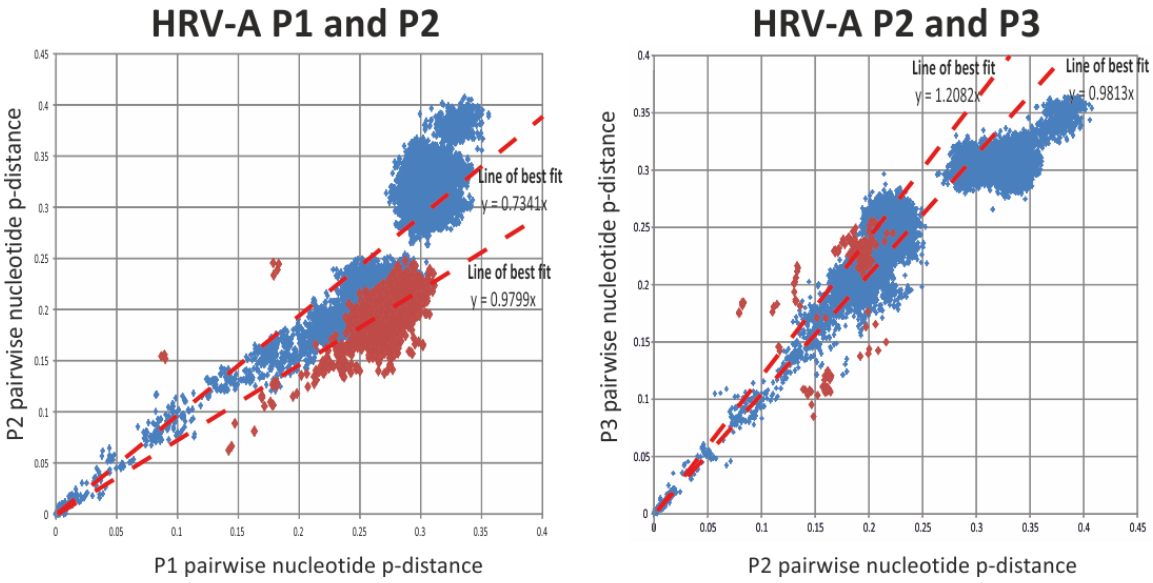


FIGURE 6.8: Comparison of pairwise nucleotide p-distances between the P1/P2 and P2/P3 regions of HRV-A. *In each distribution, the line of best fit is indicated by the red dotted line and given for putative recombinant and non-recombinant sequences. P-distance comparisons between putative recombinant sequences are shown in red.*

Within both analyses, the presence of two distinct distributions was apparent; both occurring at pairwise distances of over 0.25 in the P2 region. Detailed inspection of pairwise nucleotide p-distances indicated that these all represented p-distance comparisons between the observed clades 1 to 4 (Figure 6.5). Meanwhile, all p-distance comparisons below this threshold were between members of the same basal clades. The clear division of the two distributions further strengthens the assertion that relationships between the deep clades have not been subject to recombination within the recent history of diversification of the viruses.

6.3.7 Analysis of recombination within the coding region of HRV-B and HRV-C sequences

In contrast to the widespread phylogenetic incongruity observed within HRV-A sequences, HRV-B and HRV-C showed very limited evidence for recombination within the coding region. Phylogenetic relationships between the two regions within both HRV species were largely congruent (Figure 6.9). Both HRV-B and –C showed evidence for grouping into contiguous basal phylogenetic clades which were maintained in both regions. Due to poor basal bootstrap support, one HRV-C and seven HRV-B types were excluded from membership of any deep clade. Similar to the clustering observed in HRV-A sequences, there was only extremely limited evidence for recombination between the deep clades. The only exception was the potentially erroneous HRV-C sequence GQ223227, which changes deep clade membership between VP1 and 3Dpol.

In accordance with the observed conformity of HRV-B VP1 and 3Dpol pairwise nucleotide p-distances (Figure 6.4), the majority of HRV-B sequences displayed congruent phylogenetic relationships and maintenance of phylogenetic clustering within their types as defined by VP1 (Figure 6.9 A). In fact, only one bootstrap supported change in topology was observed between these two regions of HRV-B sequences.

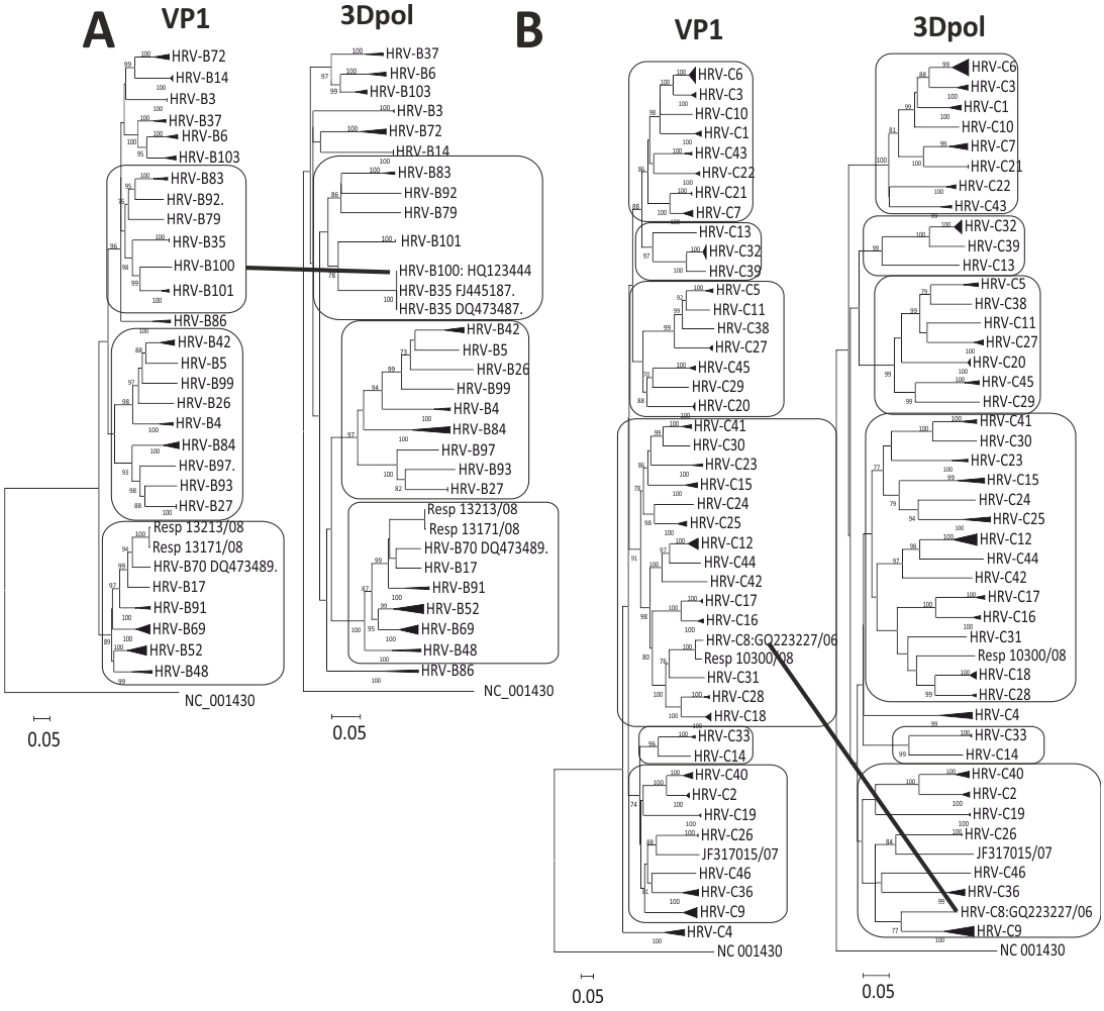


FIGURE 6.9: Neighbour joining phylogenetic trees of HRV-B (A) and HRV-C (B) VP1 and partial 3Dpol sequences. *HRV types have been collapsed for ease of reference. Putative recombinant sequences and types are marked with black lines. Contiguous clades are marked with black boxes. HRV types have retained their classical definition where we have proposed changes to these. Branches are scaled by genetic distance.*

HQ123444 is a recently published full genome sequence (Linsuwanon et al., 2011) and is currently the only representative of the HRV-B100 type. This sequence changed position to cluster extremely closely with HRV-B35 sequences in the 3Dpol region. Analysis of pairwise nucleotide p-distances revealed that this sequence had a minimum inter-type p-distance of 0.19 from HRV-B35 in VP1. However, on

inspection of HRV full genomes, this sequence is identical to HRV-B35 from position 6353, within the 3Dpol region. This observed phylogenetic incongruity was confirmed by analysis with RDP, GARD and SBP.

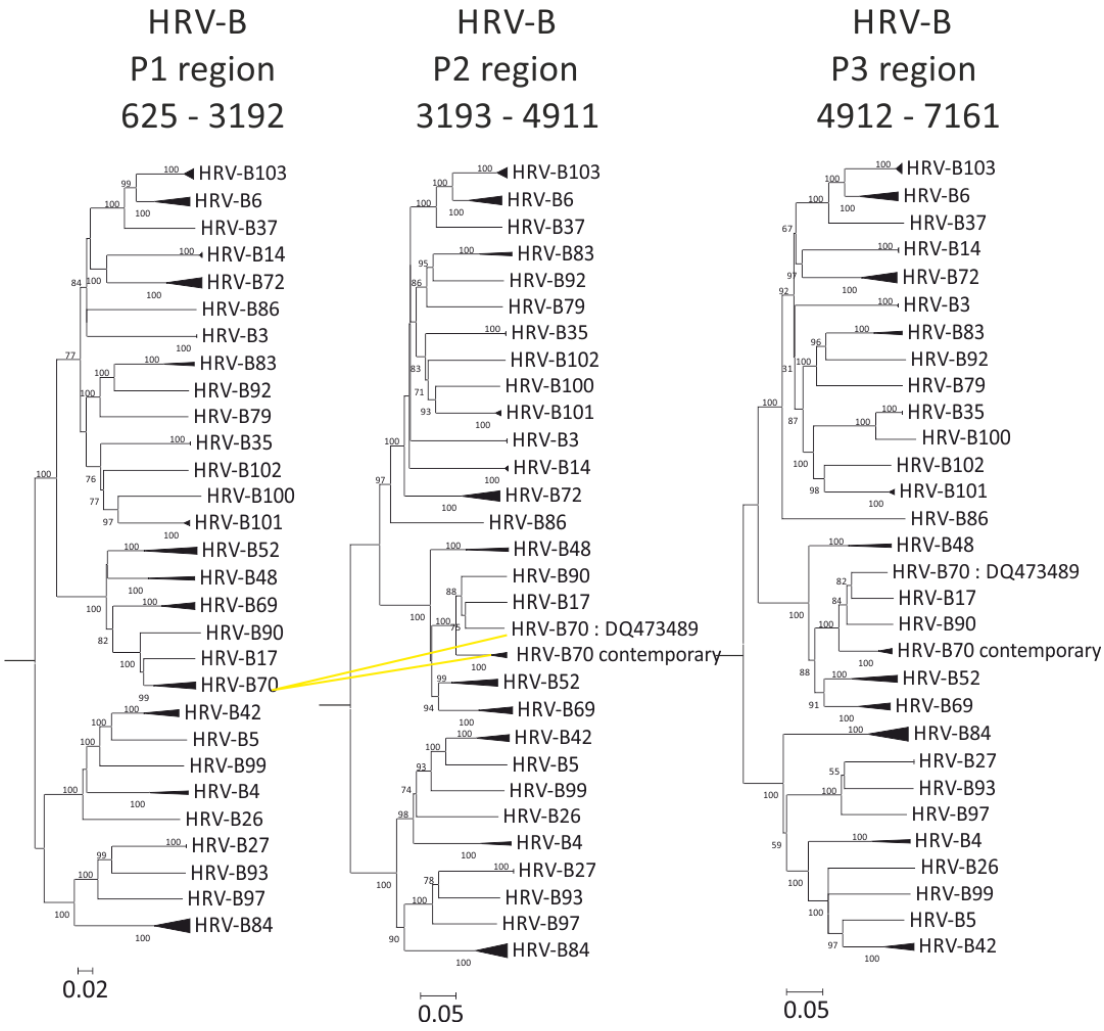


FIGURE 6.10: Neighbour joining phylogenetic trees showing the P1, P2 and P3 regions of HRV-B sequences. *P1, P2 and P3 regions are numbered by X01087: HRV-B14. One putative recombinant HRV-B type (HRV-B70) is indicated by yellow line. For ease of reference, branch to tree root has been collapsed. Branches are scaled by genetic distance.*

On inspection of phylogenetic trees constructed for the full P1, P2 and P3 regions of HRV-B full genome sequences, an incongruity in branching order within HRV-B70 was noted (shown with yellow line on Figure 6.10). Contemporary sequences which

belong to the HRV-B70 type by VP1/P1 grouping group separately from the HRV-B70 prototype strain in both P2 and P3. The prototype strain maintains its original tree position, closely related to HRV-B17. This putative recombination event was also evident on inspection of the 3Dpol trees, but only had adequate bootstrap support for the recombinant grouping when P2 and P3 trees were considered. The majority of HRV-B sequences were additionally observed to group within the three basal clades previously observed in the VP1 and 3Dpol regions (marked with black boxes on Figure 6.9).

Within HRV-C sequences, there was only one observed bootstrap supported change in tree position of a single sequence between VP1 and 3Dpol. This involved the previously reported potentially artefactually recombinant sequence GQ223227. Although this change in position was additionally observed within the P2 and P3 regions, there were no other bootstrap supported topology changes within HRV-C (data not shown). Any inconsistency in branching order showed bootstrap support below the threshold of 70% and could not be confirmed by analysis with RDP, GARD or SBP. There was also no recorded bootstrap supported change in basal clade membership between the P1, P2 and P3 regions.

6.3.8 Determining putative recombination breakpoints

In order to estimate recombination breakpoints for putative recombinant HRV sequences, all available full genomes from each noted recombinant type underwent GroupScan analysis and were compared directly to their nearest neighbours in the P1, P2 and P3 regions. Breakpoints were localised to the area where the grouping score with one control group fell and another increased. Breakpoint estimates were pinpointed by calculating the intersection of the two lines. Two representative examples of GroupScan output are given, with putative recombinant breakpoints highlighted (Figure 6.11 A).

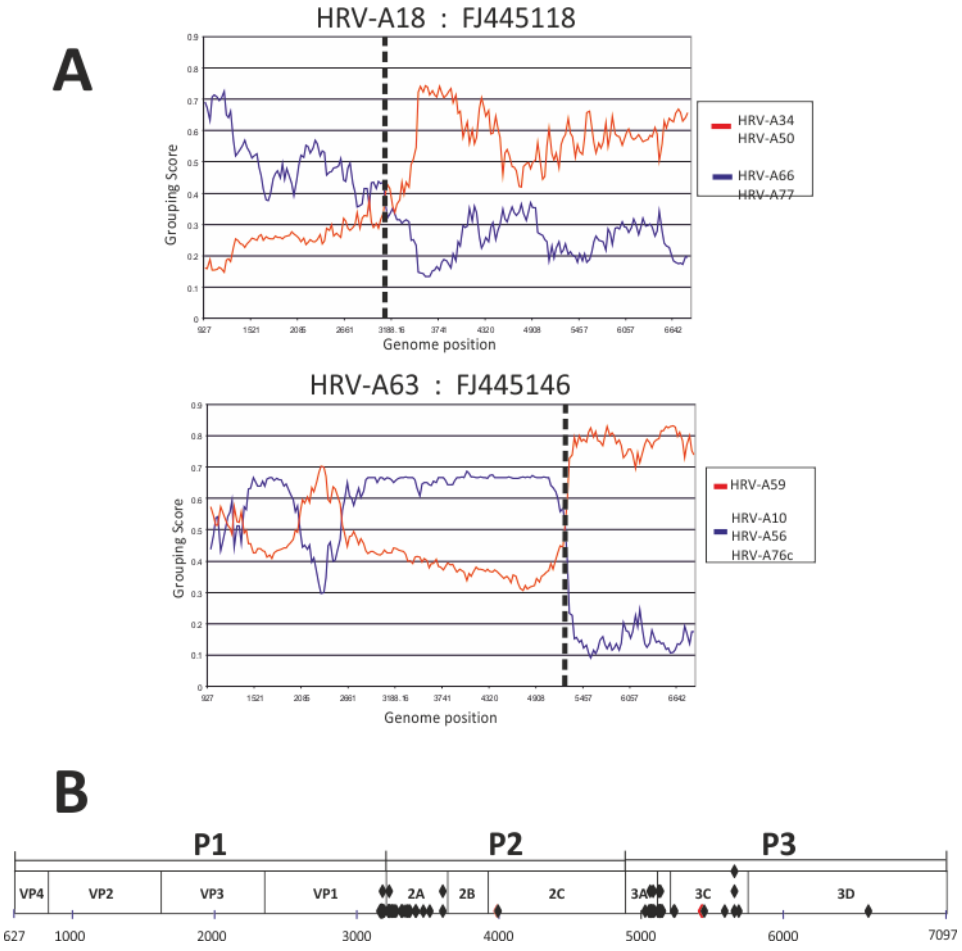


FIGURE 6.11 A: ***GroupScan of two representative recombinant HRV-A types.*** *Proposed recombination breakpoint are indicated by the dotted line.* ***B:*** ***Calculated coding region recombination breakpoints for all HRV-A recombinant types, mapped onto a diagram of the HRV-A full genome.*** *Full genome numbering is taken from FJ445111 – HRV-A1. Several sequences from large recombinant 3Dpol clade have been excluded, as complexity of nearest neighbour relationships precluded accurate estimation of recombination breakpoints.*

HRV-A18 has a breakpoint near the P1/P2 boundary which falls near position 3397 (marked by dotted line on Figure 6.11 A). Although an area of uncertain phylogenetic grouping can be observed immediately 5' and 3' to the calculated breakpoint, the only genuine point of intersection is at 3397. HRV-A63 shows a putative recombination breakpoint within the P3 region, with the intersection calculated at position 5124. The genome fragment between position 2000 and 2500

notably shows a similarly elevated grouping score with HRV-A59. However, this association only borders on significance and did not correspond to any bootstrap supported change in tree topology. It was therefore discounted.

Breakpoints were successfully determined for the majority of putative recombinant HRV-A sequences and types. However, within the large recombinant 3Dpol clade observed (Figure 6.5), several HRV-A types had very complex or closely related nearest neighbour relationships in both regions, which precluded the estimation of breakpoints. In addition, no breakpoint could be determined for the putative recombination event involving contemporary HRV-A80 or HRV-B70 sequences, as no minor recombination parent group existed within the dataset.

The majority of putative breakpoints occurred near the P1/P2 boundary, specifically within the 2A coding region (Figure 6.11 B). A second putative hotspot region was observed around the P2/P3 boundary. Most breakpoints within this region occurred within the 3A coding region or at the 3A/3B junction.

6.3.9 Recombination within the 5ˈUTR of HRV-A and HRV-B

Further to the detailed analysis of inter-species recombination within the 5ˈUTR region of HRV-A and HRV-C (presented in Chapter 5), intra-species recombination within the 5ˈUTR region of HRV-A and HRV-B was analysed. All sequences that were more than 90% complete across the 5ˈUTR fragment numbered 167 – 626 (numbered by FJ445111) were included in the analysis (167 HRV-A and 38 HRV-B) [21]. The occurrence of recombination within the 5ˈUTR was assessed by visual inspection of phylogenetic trees for HRV-A (Figure 6.12 A) and HRV-B (Figure 6.12 B) separately. An additional analysis including all three species demonstrated distinct clustering of HRV-B sequences, distant to HRV-A and HRV-C within the 5ˈUTR region, indicating no inter-species recombination involving HRV-B.

[21] A fragment significantly smaller than the full 5ˈUTR region was used for analysis, as this allowed inclusion of a much larger number of sequences.

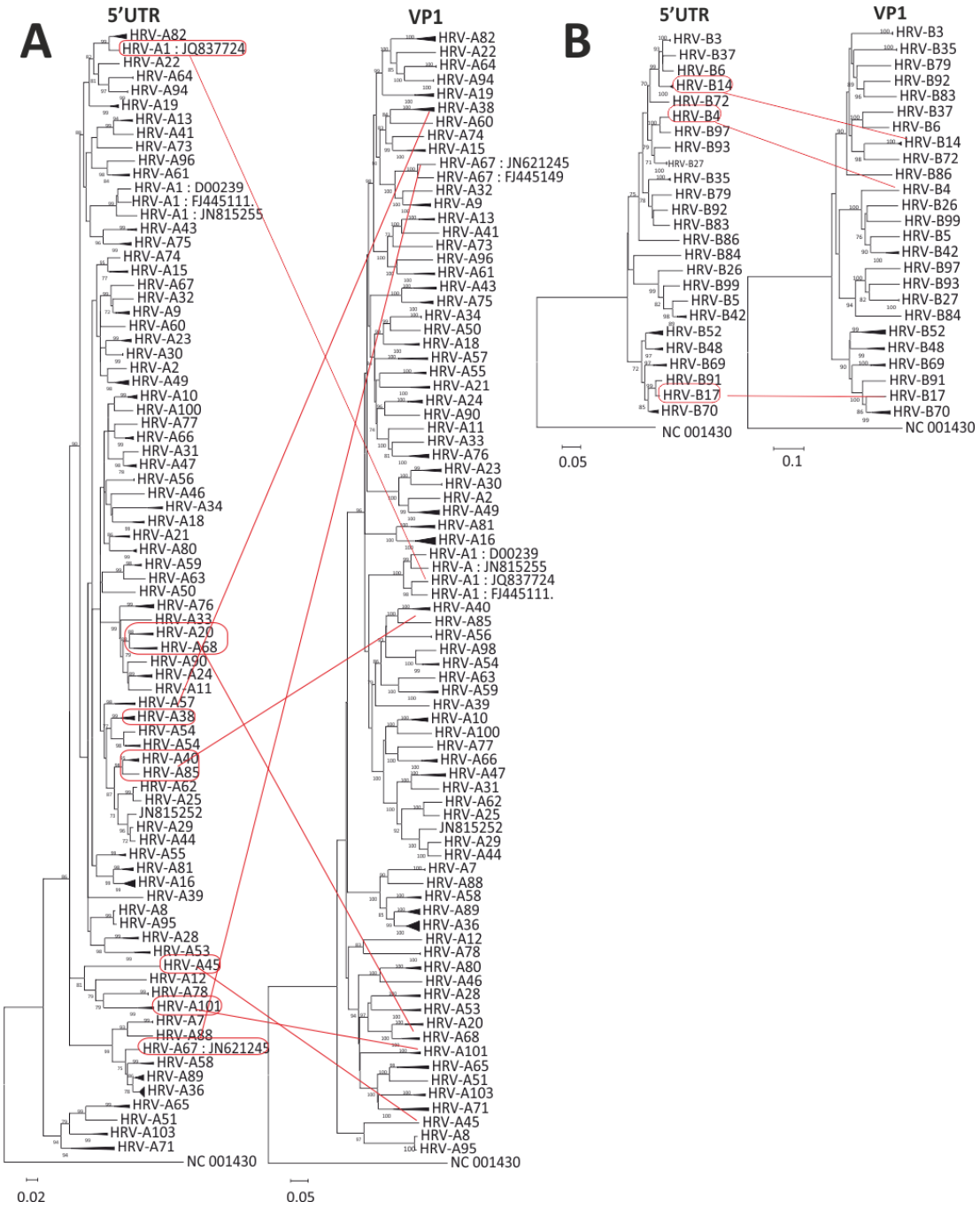


FIGURE 6.12: Neighbour joining phylogenetic trees constructed for the 5ˈUTR and VP4 regions of HRV-A and HRV-B sequences. *Putative recombinant sequences and types are marked by a red box on the 5ˈUTR tree and their corresponding position on the VP4/VP2 tree is highlighted by a red line. Branches are scaled by genetic distance.*

Within HRV-A, 17 putative recombinant sequences from nine HRV types were subject to bootstrap supported changes in nearest neighbour between the two regions (marked with black boxes on Figure 6.12). Two sequences (JQ837724 : HRV-A1 and JN621245 : HRV-A67) formed bootstrap supported groupings in the 5ˈUTR distant from other members of the same type. This may indicate the occurrence of recombination events within contemporary sequences.

In contrast, HRV-B 5ˈUTR sequences were considerably less numerous and only three putative recombinant types were noted (marked with black boxes on Figure 6.12). Of these, two consisted of a single sequence which was the sole representative of the type in question.

6.4 Discussion

The results of the analysis presented within this chapter reaffirms the assertion that recombination is a rare and sporadic occurrence in the contemporary evolution of HRV genomes. This suggests that contemporary HRV types may exist as generally segregating groups with little potential for recombination between them and that the mainstay of generation of contemporary HRV diversity is genetic drift. In contrast, a large number of recombination events were noted to have potentially occurred earlier during the period of diversification of HRV-A sequences into currently recognised types.

6.4.1 Estimation of mean substitution rates and analysis of pairwise nucleotide p-distances

This study represents the first comparison of mean substitution rates between distant genome regions of HRV-A and HRV-C. Mean substitution rates calculated for HRV-A and HRV-C correlated with published estimates for other single stranded RNA viruses (Jenkins et al., 2002; Duffy et al., 2008; Hicks and Duffy, 2011). However, most previously published studies of substitution rates with picornaviruses

focus on the capsid coding region, particularly VP1 (Calvert et al., 2010; Hicks and Duffy, 2011). One previously published estimate of the mean substitution rate within VP4/VP2 sequences of HRV-C, 6.6 x 10^{-4} substitutions/site/year, was considerably lower than our estimates for other regions of HRV-C (Briese et al., 2008). This possibly relates to the much more restricted sampling time of 32 months used in this study or to the fact that VP4/VP2 is generally a well conserved coding region with significantly less observed genetic diversity than the VP1 or 3Dpol regions.

In a published comparison of mean substitution rates within the *Picornavirus* family, members of the genus *Enterovirus* (excluded HRV species) were found to have a significantly higher mean substitution rate than other genera (including *Apthovirus, Teschovirus, Hepatovirus* and *Cardiovirus*) (Hicks and Duffy, 2011). In fact, substitution rates estimated for both the VP1 and 3Dpol regions of EV were around one order of magnitude faster than those obtained for other members of the *Picornavirus* family. Our estimates for HRV-A and HRV-C showed overlapping 95% HPD intervals with both EV and non-EV examples.

Studies of the evolution of recombinant forms of EV have included estimates of substitution rates within the 3Dpol region which are of the same order of magnitude as our estimates for HRV-A and –C (Hicks and Duffy, 2011; McWilliam Leitch et al., 2009a, 2012, 2010). In addition, when recombinant forms of EV71, E9, E11 and E30 are considered separately, the same general congruency of rates between the VP1 and 3Dpol region is observed (McWilliam Leitch et al., 2009a, 2010, 2012).

Observed similarity in mean rates of nucleotide substitution between two distant genome regions should result in a linear relationship between sequence divergence in these two regions (revealing a distribution with a gradient of 1 and a y-intercept of 0). In the resulting distribution, any discontinuities or outlying p-distance comparisons may relate to potential recombination events (coloured red and yellow

in Figure 6.4). Indeed, this relationship has been documented in EV71 isolates (McWilliam Leitch et al., 2012), where outlying data points corresponded to putative recombinant sequence groups. Within our analysis, the observed discrepancies in sequence divergence between VP1 and 3Dpol of HRV-A were typically the result of comparisons between variants that displayed bootstrap supported changes in phylogenetic clustering between the two regions. The formation of clusters of outlying data points supports the observed phylogenetic incongruity involving full HRV-A types and in particular, a large cluster of putative recombinant sequences between 0.25 and 0.34 pairwise nucleotide p-distance in VP1 (coloured red in Figure 6.4) corresponds to the widespread reorganisation of inter-type relationships observed within Clade 1 of HRV-A. Several instances of sporadic inter-type recombination observed within HRV-A are manifest as scattered discrepant comparisons (coloured yellow in Figure 6.4). In contrast, comparisons of sequence divergence in the VP1 and 3Dpol regions of HRV-B and HRV-C showed relatively few outlying data points, consistent with the observed congruence of phylogenetic relationships and generalised lack of recombination detected within these two species.

6.4.2 Analysis of recombination within the coding region of HRV-A, HRV-B and HRV-C

Despite the generally held view that HRV are not usually subject to recombination, a small number of recombination events have been hypothesized to have contributed to the formation of some recognised HRV-A and HRV-B types (Palmenberg et al., 2009). In addition, it has been speculated that the entire HRV-B species was formed by recombination between HEV and HRV-A (Tapparel et al., 2009b).

The absence of recombination documented between the two subgenomic regions at the 5' and 3' extremes of the capsid coding region (VP4/VP2 and VP1) affirms the suitability of either of these regions for the purposes of taxonomic classification and

epidemiological identification of circulating HRV types (see Chapter 4). Despite poor bootstrap support for basally branching phylogenetic relationships in both VP4/VP2 and 3Dpol, likelihood mapping analysis revealed a strong support for the representation of HRV sequences as a bifurcating phylogenetic tree in all regions studied (Figure 6.2). This led us to believe that conclusions drawn regarding potential recombination events using phylogenetic incongruity as the cornerstone of the analysis were valid. Nonetheless, supplemental examination of all putative recombinant sequences was undertaken with the panel of computational analyses described previously.

In common with previous analyses of recombination with HRV full genome sequences (Palmenberg et al., 2009) and phylogenetic incongruity between the VP4/VP2 and 3Dpol regions of HRV sequences (Savolainen et al., 2004), we found the highest proportion of recombinant sequences within HRV-A. HRV-B and HRV-C sequences showed only very limited evidence for recombination within the coding region. In contrast, Linsuwanon and co-workers found that recombination frequency within HRV full genomes was actually highest within the 3Dpol region of HRV-C (Linsuwanon et al., 2011). However, the relevance of this finding is questionable, as the study in question considered only 9 HRV-C full genome sequences, including the previously reported potentially artificially mosaic sequence GQ223227. The low sample size combined with the single aberrant sequence is likely to have given an erroneously high estimate of recombination frequency. Our similar analysis included 59 HRV-C full genome sequences and additionally excluded GQ223227, which should give a more robust estimate of recombination frequency in this species.

HRV-A sequences reliably cluster into four bootstrap supported basal clades, which were consistently observed in the regions analysed in this study (VP4/VP2, VP1 and 3Dpol). Recombination appears to have occurred relatively frequently within clades 1-3 (Figure 6.5), especially during the period of diversification of HRV-A into

currently recognised type groups. However, there was no evidence of recombination occurring between the basal clades in any coding region indicating that these clades are now potentially on distinct and diverging evolutionary paths. This concept of separate divergent evolution is supported by the inspection of comparisons of sequence divergence in the P1, P2 and P3 regions, whereby inter-clade comparisons form a discrete cluster with no overlap with intra-clade comparisons. One group of three HRV-A types (HRV-A8, -A45 and –A95) has been previously noted to be phylogenetically distinct (Palmenberg et al., 2009; Savolainen et al., 2004) and additionally possess certain RNA elements which are atypical of other HRV-A strains (Palmenberg et al., 2009). These three types have been tentatively proposed as a new species; HRV-D although they do not meet the ICTV guidelines for definition as a distinct species in terms of sequence divergence.

The most comprehensive study of recombination within all prototype full genomes of HRV previously published detected a total of 23 recombinant genomes, resulting from 12 distinct recombination events (Palmenberg et al., 2009). Notably, most HRV types within this analysis were represented by a single isolate. Our analysis uncovered putative recombination events analogous to most of the previously described events within the coding region. However, as we elected to define nearest neighbour relationships by inspection of bootstrap supported phylogenetic trees, our estimates of major and minor recombination parents occasionally differed from those reported. The detailed deep phylogeny of nearest neighbour relationships is not always well-resolved, particularly in the non-structural protein coding regions studies (P2, P3 and 3Dpol). The use of GroupScan as the mainstay of breakpoint determination allowed us to consider full clades as single nearest neighbour groups and additionally calculate breakpoints in sequences with no closely related parental sequences within the dataset. In addition to the extensive recombination involving full HRV-A types that was documented, we also recorded several instances of putative recombination among contemporary HRV-A sequences. This included

confirmation of the previously reported finding that contemporary HRV-A76 sequences have undergone a recombination event with HRV-A56 (Tapparel et al., 2009b). The original study reported a single recombinant sequence which displayed this branching pattern. However, we have noted that in fact, all contemporary HRV-A76 isolates currently known group together and therefore, that the recombinant group of HRV-A76 is the present dominant lineage. A similar pattern was also observed in several other HRV types, indicating that while contemporary recombination occurring with the 50 year time span of this dataset is not frequent, it can occur.

In contrast, only the previously reported putative recombinant HQ123444 (Linsuwanon et al., 2011) was noted in analysis of HRV-B sequences. Due to differing analysis methods used, the previous study did not detect the 100% identity of this sequence with prototype HRV-B35 strains from position 6353 onwards and instead noted an association with HRV-B35 throughout the length of the genome. As there are no contemporary HRV-B35 sequences available and HQ123444 is the only currently known example of HRV-B100, it is not possible to speculate on the validity of this observation. It is notable, however, that the breakpoint noted here is near the same region as the previously described putative sequencing error observed for GQ223227 (HRV-C8) (McIntyre et al., 2010). It is possible that this putative recombinant sequence may represent a sequencing or assembly error.

In common with earlier work on recombination within the coding region of HRV-C, only one recombinant sequence was observed within this species. However, although the renewed analysis included HRV-C strains spanning a 14 year time period, it is regrettably still limited in its scope by the lack of older strains. Notably, all instances of recombination involving only contemporary sequences of HRV-A and HRV-B documented involved a sequenced example from the collection of prototype strains lodged at the ATCC that were originally collected over 40 years

ago. It is not beyond the realm of possibility that HRV-C strains undergo recombination among contemporary sequences at similar frequencies to HRV-A, but the time span of the current data set renders detection impossible.

6.4.3 Putative recombination breakpoints within the coding region

Recombination breakpoints were determined for all putative recombinant HRV-A sequences with adequately supported nearest neighbour groups (Table 6.1). For putative recombination events involving very closely related HRV types (for instance, HRV-B17 and –B70), recombination breakpoints could not be accurately determined.

Similarly to documented observations within EV (Lindberg et al., 2003; Lukashev et al., 2005) and human parechovirus (Benschop et al., 2010b), the majority of recombination breakpoints occur near the P1/P2 junction, at the boundary of structural and non-structural protein coding regions. High within-species divergence in observed within the P1 region of HRV sequences, which suggests that there may be a biological compatibility barrier which decreases the reproductive fitness of progeny which are recombinant within the capsid region. Therefore, although inter-typic recombination events likely occur in a random fashion throughout the full length of the genome in the course of a natural co-infection, recombination within the capsid region may not be represented in routine clinical sampling which only detects the predominant variant.

6.4.4 Recombination within the 5ˈUTR of HRV-A and HRV-B

Despite the lack of recombination observed within the coding region of HRV sequences, there were several documented instances of phylogenetic incongruity within the 5ˈUTR region of HRV-A and HRV-B. Although inter-species recombination has been documented among EV isolates (Santti et al., 1999)**,** the only instances observed within HRV sequences involved the previously documented

relationship between HRV-A and HRV-C (McIntyre et al., 2010; Huang et al., 2009; Wisdom et al., 2009a). Analysis restricted to HRV-A and HRV-B alone showed total segregation of the two species in the 5ˈUTR, as in all other gene regions.

In HRV-A, basal clades 1 to 4 were not maintained in the 5ˈUTR region. We noted considerably more putative recombination events within the 5ˈUTR of HRV-A than previous studies (Palmenberg et al., 2009) and these included two instances of contemporary recombination involving single sequences grouping separately from other sequences of the same type. The discrepancy in these analyses is potentially explained by the differing alignment methods employed and by the use of phylogenetic incongruence as the mainstay of our recombination analysis. In fact, several putative recombinant sequences that were previously noted by Palmenberg and co-workers displayed inadequate bootstrap support on phylogenetic trees to be considered as recombinant by our methods (Figure 6.12). This lack of support for previously reported recombinant sequences was especially evident upon analysis of HRV-B. However, the 5ˈUTR region of HRV-B has not been extensively sequenced to date and currently a mere nine HRV-B types are represented by more than one sequence. Future studies may benefit from a focus on obtaining sequence data in order to perform a more comprehensive analysis of this region.

6.4.5 Conclusion

This study demonstrates that, despite a close phylogenetic relationship with EV, recombination within the history of diversification of all three species of HRV was substantially less frequent than that observed within EV species. Following upon previous reports which have provided some glimpses of the occurrence of recombination in HRV-A (Palmenberg et al., 2009; Tapparel et al., 2009b; Savolainen et al., 2004), our study represents a comprehensive analysis of recombination in all published HRV full genome sequences to date, an additional analysis of HRV sequence fragments obtained from isolates spanning a 14 year time

period and, to our knowledge, the first attempt to catalogue all currently detectable recombination events and potential breakpoints within HRV-A.

Chapter 7
Concluding remarks

7.1 A broad view of the global importance of HRV

Acute respiratory infections are one of the leading causes of paediatric mortality worldwide (Bryce et al., 2005). Severe acute respiratory diseases account for around 20% of deaths in children under 5 (in the 42 countries where 90% of global child deaths take place) (Black et al., 2003; Bryce et al., 2005). This figure rises to 29% when pneumonia and sepsis in neonates is additionally considered (Bryce et al., 2005). Although more widely feared serious respiratory infections (including influenza, tuberculosis and *Streptococcus Pneumoniae*) are responsible for by far the largest proportion of these illnesses and deaths, the role of HRV as a potential cause of lower respiratory tract disease, severe bronchiolitis and life-threatening exacerbations of chronic lung disease is now clear (Papadopoulos, 2002; Renwick et al., 2007; Broberg et al., 2011; Mallia et al., 2011; Bizzintino et al., 2011; McManus et al., 2008).

The predominance of the life-threatening infections mentioned above and opportunistic infections secondary to human immunodeficiency virus (HIV) infection in the developing world has generally prevented most investigations of seemingly trivial pathogens, such as HRV and as such, there is only very little data available on the regional circulation and clinical impact of HRV (Smuts et al., 2011; Chidlow et al., 2012). However, chronic diseases such as chronic obstructive pulmonary disease (COPD) and asthma, traditionally the scourges of affluent western societies, are now known to be making significant headway into these regions (Chan-Yeung et al., 2004). In addition, understanding the circulation, epidemiology and evolution of HRV can theoretically inform the development of targeted therapeutic strategies, thereby reducing not only the clinical burden of these infections but the

inappropriate prescription of antibiotics for viral respiratory tract illness. The common cold is one of the most common reasons for inappropriate antibiotic prescription (Nyquist et al., 1998) and with the threat of widespread antibiotic resistant pathogens looming, the urgency of dealing with this situation on a global scale has become clear. The realisation of the potential importance of these pathogens has meant that the scope of research interest in HRV has been increasing steadily over the last 20 years and over 200 research papers were indexed on PubMed in 2012 (www.ncbi.nlm.nih.gov/pubmed).

7.2 Novel findings in the clinical epidemiology of HRV

One of the main reasons for this substantial upswing in HRV research globally has been the realisation that these viruses are potentially responsible for more severe clinical illnesses than was previously thought. Most of the published analyses of the impact of HRV in severe illness have focussed on both severe respiratory tract infections (Broberg et al., 2011; Renwick et al., 2007; Fuji et al., 2011; Hicks et al., 2006) and occasional reports of systemic disease in young children (Tapparel et al., 2009c; Broberg et al., 2011). The results of epidemiological studies undertaken during the course of this PhD (Chapter 3) suggested a surprising link between HRV isolation from stool and gastrointestinal infections and also a potential role for HRV in aseptic meningitis (Harvala et al., 2012b). Since the publication of our findings, two other studies have also been published that report detection of HRV RNA in stool samples (Lau et al., 2012; Honkanen et al., 2013).

The closely related human EV are routinely isolated from both human stool samples obtained in cases of clinical illness and in environmental surveillance of sewage (Benschop et al., 2010a). In fact, screening sewage samples is one of the mainstays of the surveillance aspect of the poliovirus eradication campaign. However, HRV are generally thought to be acid labile and therefore unable to withstand the highly

acidic environment of the gastrointestinal tract. Exposure to acidic vapour was reported to cause conformational changes in the capsid structure of HRV-B14 and these changes may prevent uncoating of the RNA genome (Giranda et al., 1992). Although the acid sensitivity of several HRV strains has been demonstrated experimentally, the generally held wisdom that HRV are almost universally acid labile has been called into question several times. Several acid stable mutants of HRV were reported over 20 years ago (Skern et al., 1991) and viable HRV have been isolated from sewage samples by cell culture based screening techniques (Blomqvist et al., 2009). The fact that the study that found HRV in sewage recorded only 24 strains over an 8 year period is likely to be secondary to the usage of cell culture based methods that are specific for the isolation of human EV. More recent studies, including ours, undertaken using molecular methods and specifically screening patients with gastrointestinal symptoms have found rates of HRV detection in stool ranging from 8-10% (Honkanen et al., 2013; Harvala et al., 2012b; Lau et al., 2012). Not only were the HRV detected in sewage able to propagate *in vitro*, but actually one clinical HRV isolate from a stool sample was subsequently grown in RD cells (Honkanen et al., 2013; Blomqvist et al., 2009). This indicates the surprising presence of viable and infectious HRV in both sewage and stool samples.

Interestingly, no study has found evidence for a distinct genetic subgroup of HRV being more commonly isolated from the gastrointestinal tract. Instead, in accordance with our findings, studies have generally reported a high level of similarity with respiratory strains. These findings indicate that not only are HRV potentially able to withstand more acidic environments than was previously thought, but the boundary between the clinical manifestations of HRV and EV infections may be more blurry than it is traditionally considered. In addition to the potential link between HRV and gastrointestinal infections, certain human EV strains, such as EV-D68 and EV-C109 are also associated with URTI. (Oberste et al., 2004b; Yozwiak et al., 2010) Future studies in this area could comprise of the direct experimental evaluation of infectivity

of acid exposed HRV isolated from both the gut and respiratory tract. The use of a control group of known acid-stable EV would allow direct comparison between the two groups of viruses.

The distribution of ages of patients with HRV detected in their stool samples included mainly young children and the elderly. Both of the other published studies screened samples from children alone (Honkanen et al., 2013; Lau et al., 2012) and so there is a striking lack of data regarding possible HRV detection in the stool of adult patients. There was also no detailed information available as to the underlying medical history of patients with HRV detected in their stool samples and therefore pre-existing gastrointestinal pathology and drug history was unknown. In future, prospective studies conducted should probably seek to include this information, especially where concerning the presence of gastro-oesophageal reflux disease (GORD) and subsequent usage of over-the-counter remedies for this condition. Commonly prescribed treatments for gastro-oesophageal reflux are generally aimed at reducing gastric acidity. Drugs such as Omeprazole (proton pump inhibitors; PPIs) are extremely commonly prescribed both as direct treatment for GORD and as a supplementary gastro-protective agent in cases of therapy with certain painkillers within the class of non-steroidal anti-inflammatory drugs (NSAIDs), very commonly used for the treatment of arthritis. PPIs are also, rather controversially, used occasionally to treat GORD in infants and young children. Speculatively, there may be potential for a link between the usage of these medications, subsequent alteration of the pH of gastric contents and infection with HRV. The collection of detailed medical and prescription histories in a prospective study would allow examination of these factors.

A second surprising finding was the amplification of HRV RNA from the CSF of a systemically unwell infant. HRV has not, to our knowledge, been documented as a potential cause of aseptic meningitis. The finding does suggest that under certain

circumstances, HRV is able to traverse the blood brain barrier and therefore potentially cause an infection of the CNS. The fact that this appears to have been a comparatively rare occurrence, as compared to CNS infection with the highly genetically and structurally similar human EV may in theory be partially explained by the apparently limited tendency of HRV to cause a systemic viraemia. This is thought to be a vital step in the spread of EV infections to the CNS, except in the case of PV, which are hypothesized to be additionally able to directly invade peripheral nerves (Pallansch and Roos, 2007). In two studies of HRV viraemia, only 11-12% of symptomatic HRV infections were linked with HRV detection in blood samples and this detection frequency was reduced to zero when those without respiratory symptoms were considered (Xatzipsalti et al., 2005; Fuji et al., 2011). However, studies on the rates of HRV viraemia in patients with symptomatic gastrointestinal infections and HRV detected in stool have not, to our knowledge, been undertaken. Studies such as this would provide further clues as to why, when HRV appears to be able to both replicate in the gastrointestinal tract and cross the blood brain barrier, it does not cause a similar proportion of aseptic meningitis as human EV.

7.3 Classification of HRV into genotypically defined types

A core component of the work presented within this thesis has been the development of a system of defining and classifying HRV on the basis of readily available sequence data. The determination of associations of specific HRV types with certain clinical illnesses, if they exist, is an issue of great importance; especially with the hypothesis that certain HRV strains may be more likely to both be implicated in the development and cause severe exacerbations of asthma (Palmenberg et al., 2010). Consistency in classification and nomenclature of HRV types is essential to these endeavours.

The system of classification for all three species of HRV that we have proposed is similar in nature to that currently in use for the definition of new EV types (Oberste et al., 1999a, 1999b). The realisation that the majority of HRV-C sequences are recombinant in the 5ˈUTR region (Huang et al., 2009; Wisdom et al., 2009a; McIntyre et al., 2010) should preclude the use of this region for this purpose as the majority of HRV-C strains will be misidentified. The use of capsid regions only for the purposes of HRV typing should not cause underestimation of the importance of non-structural protein coding regions for investigation of viral phenotypes. For example, one study of EV-A71 pathogenesis in mice found specific mutations in the 3Dpol region of strains with reduced virulence (Chang et al., 2010). The extensive recombination observed within EV species means that typing EV by VP1 alone is tantamount to knowing that around one third of the genome is closely related to a known prototype strain. If virulence and tropism influencing residues are situated outside the capsid, then a more comprehensive approach to EV typing could be warranted. Whole genome sequencing has been suggested as the most appropriate method for HRV typing (Palmenberg et al., 2009). However, as recombination within the coding region appears to occur only very infrequently within the contemporary evolution of HRV, this approach may not be necessary.

We also noted that the three species of HRV appear to harbour considerable antigenic diversity both within and between species. Over 140 HRV types over three species are recognised compared to 130 types (in nine species) for EV and 11 (in four species) for *Apthovirus* (Table 1.1). This discrepancy may be due to the ubiquitous nature and constant circulation of HRV, meaning that many more types are detected and catalogued in a single study than other picornavirus groups. The action of diversifying selection pressure on residues within the capsid coding region that specifically coincide with neutralizing antigenic sites has previously been described (Kistler et al., 2007b) and one could speculate that such pressure could lead to the routine emergence and identification of many new HRV types. However,

a large number of the HRV types that have been newly discovered in recent years have shown evidence of circulation long before their initial discovery (Briese et al., 2008). These cannot be properly called emerging viruses but rather new discoveries made possible by the increasing scope of molecular detection methods. In addition, evidence of the contemporary diversification of HRV into genotypically distinct types was only rarely evident within our dataset, although several divergent strains of a few types (including HRV-A29) were documented (Chapter 4). These may represent the gradual creation of antigenically distinct types. Further studies focussing on exploring serological cross-reactivity between members of genotypically defined types in all three species would be beneficial to gaining a deeper understanding of these relationships.

7.4 Recombination in the evolution of HRV

The process of recombination can generate strains with differing virulence patterns and can cause reversion in the attenuated PV vaccine (Macadam et al., 1989; van der Sanden et al., 2011; Burns et al., 2013). The circulation of recombinant strains has been frequently reported in many other picornaviruses, including human EV (Oprisan et al., 2002; Simmonds and Welch, 2006; Bouslama et al., 2007; McWilliam Leitch et al., 2010; Santti et al., 1999; McWilliam Leitch et al., 2012; Oberste et al., 2004d, 2004a; Lindberg et al., 2003; Andersson et al., 2002), *Apthovirus* (Heath et al., 2006; Simmonds, 2006), *Cardiovirus* (Drexler et al., 2010; Blinkova et al., 2009) and *Parechovirus* (Benschop et al., 2008; Calvert et al., 2010). In fact, the occurrence of phylogenetic incongruence between the structural and non-structural genome regions of many picornaviruses has led to a theory of separate and modular evolution of these two genome regions (Lukashev et al., 2005; Santti et al., 1999). In contrast, the studies contained within this thesis (Chapters 5 and 6) have generally supported a low level of detectable recombination within the coding region of HRV genomes (McIntyre et al., 2010, 2013b). In particular, there was little

evidence within HRV for the widespread recombination between contemporary strains observed in EV. Recombinant strains of EV-B have been documented to circulate in outbreaks, where one recombinant form (a combination of a particular capsid and non-structural region) rises to dominate circulation for a few years and then disappears entirely (McWilliam Leitch et al., 2010, 2009a). In contrast, the infecting HRV population appears to consist of a great number of types which circulate almost continuously. These types appear to be somewhat stable and show congruent within-type phylogeny between their structural and non-structural genome regions. As EV and HRV in particular exhibit a highly similar genome structure and a degree of sequence homology, the mechanism by which recombination is restricted in HRV remains puzzling.

Although the occurrence of recombination during genome replication is essentially a random process and probably occurs rather frequently, recombinant genomes must necessarily compete with parental strains in order to come to form a detectable fraction of the population. The fact that recombinant HRV genomes are so rarely observed within contemporary datasets suggests that recombinant genomes may be generally less fit than parental strains. The fitness of the generated virus depends not just on the functioning of individual encoded proteins but on the interactions between different parts of the genome. For instance, the interaction between the 3CD protein and the 5ˈUTR is an essential component of negative strand synthesis and therefore RNA replication (Rohll et al., 1994; Andino et al., 1990) . These interactions have likely been somewhat optimised over time by co-evolution of different genome segments and therefore, even biologically viable recombinants are likely to exhibit decreased fitness when compared with parental strains. A similar point has been demonstrated by *in vitro* studies of recombination in the 5ˈUTR region of HRV. Artificial chimaeras were created between the coding genome of HRV-A16 and the 5ˈUTR region of various different members of the *Enterovirus* genus (comprising of an HRV-Ca, HRV-Cc, HRV-B and human EV strain). Interestingly, the study found

that each chimaera was viable and could replicate efficiently in cell culture but in competition experiments was consistently less fit than the wild type virus (Schibler et al., 2012). This suggests that while certain parts of the genome may be functionally interchangeable, the restricting factor in the detection of recombinant genomes may be a decline in viral fitness. Indeed, this functional compatibility within the 5ˈUTR region can extend to highly divergent viruses. A study which replaced the IRES of a PV1 virus with that of EMCV found that this still resulted in the creation of a viable chimaera (Alexander et al., 1994). This finding is even more remarkable when considered in light of the fact that this indicates a degree of functional compatibility between a Type I and Type II IRES, which have little or no homology in either sequence or structure. Notably there is also evidence for the exchange of IRES elements between diverse picornavirus genera. For example, as previously discussed the Type IV (pestivirus related) IRES has been found in both *Sapelovirus* and *Avihepatovirus*. The extremely rare isolation of recombinants from such divergent viruses in nature may be secondary to decreased evolutionary fitness or to their limited opportunity to co-infect the same host cell.

Although the well-conserved 5ˈUTR region appears to be somewhat interchangeable between diverse viruses, the possibility of a biological compatibility barrier between the more divergent structural and non-structural regions of divergent viruses has been previously discussed (Kistler et al., 2007b; Simmonds and Welch, 2006; Lukashev et al., 2005). The theory states that viruses that bypass a particular divergence threshold in the non-structural region may lack the ability to form viable recombinants, as vital interactions between different encoded proteins may be severely disturbed. HRV are generally highly genetically diverse within and between species in both structural and non-structural genome regions (Kistler et al., 2007b) and we have demonstrated a generally higher within-species divergence for HRV-C than other HRV (Chapter 5). This co-incides with the almost completely absence of recombinant HRV-C genomes. In contrast, the more frequently

recombinant EV-B shows considerably decreased divergence in the non-structural protein region (Lindberg et al., 2003). This may allow different genome segments to have a higher degree of functional compatibility and therefore permit successful competition of recombinant genomes with parental strains. It has been suggested that G+C content may play a role in the frequency of generation of recombinant genomes; whereby high G+C content increases recombination frequency by 1.3 times and the existence of GC tracts increase it even further (Runckel et al., 2013). This study used a deep sequencing technique to obtain sequences from all PV genomes generated during the replication cycle and so recovered sequences were not restricted by the viability or fitness of the variant. If this observation can be assumed to be generally applicable to HRV as well as PV, it may suggest that the frequency of generation of recombinants in HRV during replication is less than in EV secondary to a generally lower G+C content. This would also suggest that recombination occurs slightly more frequently in the more G+C rich HRV-C genomes than other species of HRV. However, the lack of observation of such recombinants suggests that certain restrictions on biological viability and evolutionary fitness render these recombinants less able to compete with wild type viruses.

The emergence of epidemic strains secondary to recombination of the 5'UTR region in EV has been documented (van der Sanden et al., 2011). In addition, strains of EV-A71 with altered virulence secondary to mutations in the 3Dpol region have been noted (Chang et al., 2010). EV-B circulates in a cyclical fashion, whereby recombinant strains predominate for a certain period of time and then are periodically replaced. This constant shuffling of distinct genome modules in EV may potentially contribute to the epidemic circulation of certain strains. Studies of HRV epidemiology, such as those outlined in Chapter 3, generally note a lack of predominance of a certain HRV species or strain circulating in a particular time scale. The lack of recombination and subsequent modular evolution of structural and

non-structural genome regions in HRV may be a partial explanation for the patterns of circulation observed in these viruses.

7.5 Conclusion

In summary, the work undertaken throughout the course of this PhD has focussed on the epidemiology, classification and evolution of HRV. In addition to the discovery of a novel potential association between HRV infection and both gastrointestinal and CNS disease (Chapter 3), we have developed proposals for the classification of HRV into genotypically defined types and applied these guidelines to the classification of all available HRV sequence data into over 140 types (Chapter 4). The classification of all HRV sequences in this manner allowed investigation of the occurrence of recombination in the history of diversification of all three species of HRV (Chapter 5 and 6). We found only limited evidence of recombination occurring within the contemporary diversification of HRV types, as evidenced by the congruent phylogenetic groupings of sequences. However, we did find evidence for phylogenetic incongruity affecting entire HRV-A types, suggesting that these viruses may have been subject to extensive recombination during the period of their diversification into distinct HRV types.

One potential long-term application of research into the genetic diversity of HRV and evolutionary processes shaping this diversity is the potential identification of genome regions which may serve as therapeutic targets for the development of both antiviral drugs and vaccines. The considerable genetic diversity of HRV renders the formulation of a common vaccine for even a single HRV species extremely challenging. However, the taxonomic identification and classification of infecting strains may allow future determination of potential associations with certain strains and severe diseases. Studies of evolution and recombination in HRV may reveal certain exploitable similarities between related HRV types. For instance, the finding

that antibodies to the well-conserved VP4 region may exhibit cross-serotypic neutralisation properties (Katpally et al., 2009) coupled with our observations that the VP4 region of HRV is both well conserved and largely non-recombinant throughout all three species may have implications for the development of such targeted therapies. In addition, the finding that the 2A protein is very well conserved within HRV-A and HRV-C may highlight this region as a particularly attractive therapeutic target. With global collaboration and studies concentrated on aspects of epidemiology and evolution of HRV strains, it may actually be possible in the future to make an impact on the considerable clinical and economic burden of these extremely common pathogens.

Appendix 1 – List of Primers

Name of primer set	HRV-Species	Orient[a]	Primer Number[b]	Numbered by:	Sequence
Primers for screening (5'UTR) for and typing (VP4/VP2) all HRV					
UTR Screen	All HRV	OS	178	B14 :NC_001490	HCAAGYACTTCTGTYWCCCCSG
UTR Screen	All HRV	OAS	573	B14 :NC_001490	GAAACACGGACACCCAAAGTAGT
UTR Screen	HRV-A/-B	IS	367	B14 :NC_001490	CYAGCCTGCGTGGCKGCCWRC
UTR Screen	HRV-C	IS	367	B14 :NC_001490	GTAGCCYGCGTGGTGCCCWGC
UTR Screen	All HRV	IAS	477	B14 :NC_001490	TTAGCCRCATTCAGGGGCCGG
VP4/VP2	All HRV	OS	445	C4:EF582385	CCGGCCCCTGAATGYGGCTAA
VP4/VP2	All HRV	IS	533	C4:EF582385	ACCRACTACTTTGGGTGTCCGTG
VP4/VP2	All HRV	IAS	1066	C4:EF582385	TCWGGHARYTTCCAMCACCANCC
VP4/VP2	All HRV	OAS	1104	C4:EF582385	ACATRTTYTSNCCAAANAYDCCCAT
Primers for the amplification of the 5'UTR of HRV-C					
UTR 1	HRV-A	OS/IS	1	C4:EF582385	TTMAAACAGCSGBNNGGTTGYACCCA
UTR 1	HRV-C	OS/IS	1	C4:EF582385	TTMAAACTGRRWBYRGGTTGYTCCCA
UTR 1	HRV-A/C	OAS	375	C4:EF582385	GCAGGSMRCCACGCRGGCT
UTR 1	HRV-A	OAS	375	C4:EF582385	GCAGGCAGCCACGCAGGCT
UTR 1	HRV-C	OAS	375	C4:EF582385	GCAGGGCACCACGCGGGCT
UTR 1	HRV-A	IAS	355	C4:EF582385	GAHCACTGTYRCCAGTGGGG
UTR 1	HRV-C	IAS	354	C4:EF582385	GACACCGTCGCCGGTGGGG
UTR 2	HRV-A/C	S	164	C4:EF582385	HCAAGYACTTCTGTYWCCCCSG
UTR 2	HRV-A/C	OAS	560	C4:EF582385	RGAAACACGGACACCCAAAGTAGT
UTR 2	HRV-A/C	IAS	550	C4:EF582385	ACACCCAAAGTAGTYGGTYCCR
UTR 3	HRV-A/C	A	445	C4:EF582385	CCGGCCCCTGAATGYGGCTAA

Name of primer set	HRV-Species	Orient	Primer Number	Numbered by:	Sequence
UTR 3	HRV-A/C	OAS	1066	C4:EF582385	TCWGGHARYTTCCAMCACCANCC
UTR 3	HRV-A/C	IAS	868	C4:EF582385	ATAGTRATTTGYTTDAGCCTATCDGAV A
Primers for the amplification of the VP1 region of HRV-A, -B and –C					
A VP1	HRV-A	OS	1943	A1:D00239	MGHTTYAGYTTYATGTTYTGTGG
A VP1	HRV-A	OAS	3485	A1:D00239	CCACARTCWCCWGGYTCACADGG
A VP1	HRV-A	IS	2418	A1:D00239	TRGAYGCWGCWGARACWGG
A VP1	HRV-A	IAS	3290	A1:D00239	GTRTTTGTKCGGTADATGAYTARRTC
A VP1	HRV-A	IAS	2504	A1:D00239	CTYTCHAYRCTCATYTCATC
A28 VP1	HRV-A28	IAS	2702	A1:D00239	ACTTGCAACTGACACCACTAT
B VP1	HRV-B	OS/IS	2250	B14:X01087	TAGTGCRTGTCCAGATTTYA
B VP1	HRV-B	OAS	3455A	B14:X01087	TAACTACCACCTTCAATCCA
B VP1	HRV-B	OAS	3455B	B14:X01087	TAGTTTCCTCCTTCAATCCA
B VP1	HRV-B	IAS	3256	B14:X01087	ARRTGRTCYTCHTGYGTCAT
C VP1 PCR 1	HRV-C	OS	1986	C4:EF582385	ATRGCMTACACMCCHCCWGG
C VP1 PCR 1	HRV-C	OAS	3449	C4:EF582385	YTTVCCWCCACARTCHCCWGG
C VP1 PCR 1	HRV-C	IS	2256	C4:EF582385	ATGHTRCGWGAYACWCCHATGAT
C VP1 PCR 1	HRV-C	IASa	2852	C4:EF582385	ATCATAHCCATCATARAAVATGTARTA
C VP1 PCR 1	HRV-C	IASb	2852	C4:EF582385	ATCATAHCCATCATARAAVACATARTA
C VP1 PCR 1	HRV-C	IS	2745	C4:EF582385	GCYTCHAACCCHAGTGTNTTYT
C VP1 PCR 1	HRV-C	IAS	3151	C4:EF582385	GTRTGNACAWAHAKRTCACTRGGTCC
C VP1 PCR 2	HRV-C	OAS	3882	C4:EF582385	ARBCCHYKDGCDGCRTTRCA
Primers for the amplification of the 2A region of HRV-C					
C2A	HRV-C	OS	2832	C4:EF582385	GCHTAYTAYATKTTYTATGATG
C2A	HRV-C	OAS	3882	C4:EF582385	ARBCCHYKDGCDGCRTTRCA
C2A PCR1	HRV-C	IS	2898	C4:EF582385	AATGAYAYGGGWWSASMYTRTG
C2A PCR 1	HRV-C	IAS	3429	C4:EF582385	YTTVCCWCCRCARTCHCCHGG
C2A PCR 2	HRV-C	IS	3126	C4:EF582385	GGNCCHAGTGAYMWVTWTG
Primers for the amplification of the 3Dpol region of HRV-A, -B and –C					
A 3Dpol	HRV-A	OS	5657	A1:D00239	AAYGGNARSAHGGNTTYKCHGC
A 3Dpol	HRV-A	OAS/IAS	7130	A1:D00239	ARAAYTTBTCRTACCAYTCRTG

Name of primer set	**HRV-Species**	**Orient**	**Primer Number**	**Numbered by:**	**Sequence**
A 3Dpol	HRV-A	IS	6464	A1:D00239	TTYGAYTAYWCBAAYTATGAYGG
A 3Dpol	HRV-A	OS	6323	A1:D00239	TVAATGAYTCWGTNAAYATGAG
B 3Dpol	HRV-B	OAS	7030	B14:X01087	CCWGARTGCCAGGCYADCATRCA
B 3Dpol	HRV-B	IS	6475	B14:X01087	TTYGAYTAYWCHAAYTWYGATGC
B 3Dpol	HRV-B	IAS	6974A	B14:X01087	TCVTRDATDTCHHGCATYGGCAT
B 3Dpol	HRV-B	IAS	6974B	B14:X01087	TCVTRDATDTCYTGCATYGGCAT
C 3Dpol	HRV-C	OS	6300	C4:EF582385	TGTGATCCWGAYRTYTTYTGGTCAC
C 3Dpol	HRV-C	OAS	6901	C4:EF582385	ACRTGRTCTTSWGTRTTCTTTGG
C 3Dpol	HRV-C	IS	6360A	C4:EF582385	TTTGAYTAYACMAAYTATGATGG
C 3Dpol	HRV-C	IS	6360B	C4:EF582385	TTTGAYTAYACMAAYTATGATGGNAG
C 3Dpol	HRV-C	IAS	6877	C4:EF582385	TCYYTDGTCCATCTRATTGAYTCA

[a] Orientation of primer sequences. OS : outer sense. OAS : outer antisense. IS : inner sense. IAS : inner antisense

[b] Primer numbering is taken from the 5' base position

Appendix 2 – Permissions for reproduction of publications and figures

Where required by the individual publishers, explicit permission for the reproduction of published articles has been sought and is included below.

SPRINGER LICENSE
TERMS AND CONDITIONS
Mar 19, 2013

This is a License Agreement between Chloe L McIntyre ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

License Number	3102590463561
License date	Mar 05, 2013
Licensed content publisher	Springer
Licensed content publication	Archives of Virology
Licensed content title	Recombination in the evolution of human rhinovirus genomes
Licensed content author	Chloe L. McIntyre
Licensed content date	Jan 1, 2013
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are the sole author of the new work
Title of your thesis / dissertation	The epidemiology, classification and evolution of human rhinoviruses
Expected completion date	Mar 2013
Estimated size(pages)	200
Total	0.00 GBP

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
Mar 19, 2013

This is a License Agreement between Chloe L McIntyre ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

License Number	3102600306975
License date	Mar 05, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Journal of Medical Virology
Licensed content title	High detection frequency and viral loads of human rhinovirus species A to C in fecal samples; diagnostic and clinical implications
Licensed copyright line	Copyright © 2012 Wiley Periodicals, Inc.
Licensed content author	H. Harvala,C.L. McIntyre,N.J. McLeish,J. Kondracka,J. Palmer,P. Molyneaux,R. Gunson,S. Bennett,K. Templeton,P. Simmonds
Licensed content date	Jan 13, 2012
Start page	536
End page	542
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Total	0.00 USD

SPRINGER LICENSE
TERMS AND CONDITIONS
Apr 30, 2013

This is a License Agreement between Chloe L McIntyre ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3138730673639
License date	Apr 30, 2013
Licensed content publisher	Springer
Licensed content publication	Cellular and Molecular Life Sciences
Licensed content title	The cold case: Are rhinoviruses perfectly adapted pathogens?
Licensed content author	S. Dreschers
Licensed content date	Jan 1, 2007
Volume number	64
Issue number	2
Type of Use	Thesis/Dissertation
Portion	Figures
Author of this Springer article	No
Order reference number	
Title of your thesis / dissertation	The epidemiology, classification and evolution of human rhinoviruses
Expected completion date	Apr 2013
Estimated size(pages)	200
Total	0.00 GBP

ELSEVIER LICENSE
TERMS AND CONDITIONS
Apr 30, 2013

This is a License Agreement between Chloe L McIntyre ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Chloe L McIntyre
Customer address	Infection and Immunity Division Edinburgh, EH25 9RG
License number	3138741121335
License date	Apr 30, 2013
Licensed content publisher	Elsevier
Licensed content publication	Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms
Licensed content title	Bridging IRES elements in mRNAs to the eukaryotic translation apparatus
Licensed content author	Kerry D. Fitzgerald,Bert L. Semler
Licensed content date	September–October 2009
Licensed content volume number	1789
Licensed content issue number	9–10
Number of pages	11
Start Page	518
End Page	528
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier	No

article?	
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	The epidemiology, classification and evolution of human rhinoviruses
Expected completion date	Apr 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 GBP
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 GBP

NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS
May 01, 2013

This is a License Agreement between Chloe L McIntyre ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3140160361941
License date	May 01, 2013
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Reviews Microbiology
Licensed content title	Host and virus determinants of picornavirus pathogenesis and tropism
Licensed content author	J. Lindsay Whitton,Christopher T. CornellandRalph Feuer
Licensed content date	Oct 1, 2005
Volume number	3
Issue number	10
Type of Use	reuse in a thesis/dissertation

Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1 - Summary of the picornavirus life cycle
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	The epidemiology, classification and evolution of human rhinoviruses
Expected completion date	May 2013
Estimated size (number of pages)	200
Total	0.00 GBP

References

Abraham, A. A., and Holmes, W. L. (1974). Isolation and characterization of an echovirus, possible "prime" strain of echovirus type 12. *Proc Soc Exp Biol Med* 147, 444–8.

Abraham, G., and Colonno, R. J. (1984). Many rhinovirus serotypes share the same cellular receptor. *J Virol* 51, 340–5.

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., and Brown, F. (1989). The three-dimensional structure of foot-and-mouth disease virus at 2.9 A resolution. *Nature* 337, 709–16.

Agol, V. I. (1997). Recombination and Other Genomic Rearrangements in Picornaviruses. *Seminars in Virology* 8, 77–84.

Alexander, L., Lu, H. H., and Wimmer, E. (1994). Polioviruses containing picornavirus type 1 and/or type 2 internal ribosomal entry site elements: genetic hybrids and the expression of a foreign gene. *Proc Natl Acad Sci USA* 91, 1406–10.

de Almeida, M. B., Zerbinati, R. M., Tateno, A. F., Oliveira, C. M., Romão, R. M., Rodrigues, J. C., Pannuti, C. S., and da Silva Filho, L. V. F. (2010). Rhinovirus C and respiratory exacerbations in children with cystic fibrosis. *Emerg Infect Dis* 16, 996–9.

Alper, C. M., Doyle, W. J., Winther, B., and Hendley, J. O. (2008). Upper respiratory virus detection without parent-reported illness in children is virus-specific. *J Clin Virol* 43, 120–2.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–10.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–402.

Anderson, E. C., Hunt, S. L., and Jackson, R. J. (2007). Internal initiation of translation from the human rhinovirus-2 internal ribosome entry site requires the binding of Unr to two distinct sites on the 5' untranslated region. *J Gen Virol* 88, 3043–52.

Andersson, P., Edman, K., and Lindberg, A. (2002). Molecular analysis of the echovirus 18 prototype: evidence of interserotypic recombination with echovirus 9. *Virus Res* 85, 71–83.

Andino, R., Rieckhof, G. E., Trono, D., and Baltimore, D. (1990). Substitutions in the protease (3Cpro) gene of poliovirus can suppress a mutation in the 5' noncoding region. *J Virol* 64, 607–12.

Andrewes, C. H. (1961). The taxonomic position of common cold viruses and some others. *Yale J Biol Med* 34, 200–6.

Andries, K., Dewindt, B., Snoeks, J., Wouters, L., Moereels, H., Lewi, P. J., and Janssen, P. a (1990). Two groups of rhinoviruses revealed by a panel of antiviral compounds present sequence divergence and differential pathogenicity. *J Virol* 64, 1117–23.

Arden, K. E., Faux, C. E., O'Neill, N. T., McErlean, P., Nitsche, A., Lambert, S. B., Nissen, M. D., Sloots, T. P., and Mackay, I. M. (2010). Molecular characterization and distinguishing features of a novel human rhinovirus (HRV) C, HRVC-QCE, detected in children with fever, cough and wheeze during 2003. *J Clin Virol* 47, 219–23.

Arden, K., and Mackay, I. (2010). Newly identified human rhinoviruses: molecular methods heat up the cold viruses. *Rev Med Virol* 20, 156–176.

Arden, K., McErlean, P., Nissen, M., Sloots, T., and Mackay, I. (2006). Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections. *J Med Virol* 78, 1232–40.

Arola, A., Santti, J., Ruuskanen, O., Halonen, P., and Hyypiä, T. (1996). Identification of enteroviruses in clinical specimens by competitive PCR followed by genetic typing using sequence analysis. *J Clin Micro* 34, 313–8.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol Rev* 35, 235–41.

Barton, D. J., O'Donnell, B. J., and Flanegan, J. B. (2001). 5' cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. *EMBO J* 20, 1439–48.

Belsham, G. J. (2009). Divergent picornavirus IRES elements. *Virus Res* 139, 183–92.

Bennett, S., Harvala, H., Witteveldt, J., McWilliam Leitch, E., McLeish, N., Templeton, K., Gunson, R., Carman, W., and Simmonds, P. (2011). Rapid simultaneous detection of enterovirus and parechovirus RNAs in clinical samples by one-step real-time reverse transcription-PCR assay. *J Clin Micro* 49, 2620–4.

Benschop, K., Minnaar, R., Koen, G., van Eijk, H., Dijkman, K., Westerhuis, B., Molenkamp, R., and Wolthers, K. (2010a). Detection of human enterovirus and human parechovirus (HPeV) genotypes from clinical stool samples: polymerase chain reaction and direct molecular typing, culture characteristics, and serotyping. *Diagn Microbiol Infect Dis* 68, 166–73.

Benschop, K. S., de Vries, M., Minnaar, R. P., Stanway, G., van der Hoek, L., Wolthers, K. C., and Simmonds, P. (2010b). Comprehensive full-length sequence analyses of human parechoviruses: diversity and recombination. *J Gen Virol* 91, 145–54.

Benschop, K. S., Williams, C. H., Wolthers, K. C., Stanway, G., and Simmonds, P. (2008). Widespread recombination within human parechoviruses: analysis of temporal dynamics and constraints. *J Gen Virol* 89, 1030–5.

Berry, V., and Gascuel, O. (1996). On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. *Mol Biol Evol* 13, 999–1011.

Bienz, K., Egger, D., and Pfister, T. (1994). Characteristics of the poliovirus replication complex. *Arch Virol Suppl* 9, 147–57.

Bizzintino, J., Lee, W.-M., Laing, I. a, Vang, F., Pappas, T., Zhang, G., Martin, a C., Khoo, S.-K., Cox, D. W., Geelhoed, G. C., et al. (2011). Association between human rhinovirus C and severity of acute asthma in children. *Eur Resp J* 37, 1037–42.

Black, R. E., Morris, S. S., and Bryce, J. (2003). Where and why are 10 million children dying every year? *Lancet* 361, 2226–34.

Blinkova, O., Kapoor, A., Victoria, J., Jones, M., Wolfe, N., Naeem, A., Shaukat, S., Sharif, S., Alam, M. M., Angez, M., et al. (2009). Cardioviruses Are Genetically Diverse and Cause Common Enteric Infections in South Asian Children. *J Virol* 83, 4631–4641.

Blomqvist, S. (2003). Characterization of a recombinant type 3/type 2 poliovirus isolated from a healthy vaccinee and containing a chimeric capsid protein VP1. *J Gen Virol* 84, 573–580.

Blomqvist, S., Paananen, A., Savolainen-Kopra, C., Hovi, T., and Roivainen, M. (2008). Eight years of experience with molecular identification of human enteroviruses. *J Clin Micro* 46, 2410–3.

Blomqvist, S., Roivainen, M., Puhakka, T., Kleemola, M., and Hovi, T. (2002a). Virological and serological analysis of rhinovirus infections during the first two years of life in a cohort of children. *J Med Virol* 66, 263–8.

Blomqvist, S., Savolainen, C., Råman, L., Roivainen, M., and Hovi, T. (2002b). Human rhinovirus 87 and enterovirus 68 represent a unique serotype with rhinovirus and enterovirus features. *J Clin Micro* 40, 4218–23.

Blomqvist, S., Savolainen-Kopra, C., Paananen, A., Hovi, T., and Roivainen, M. (2009). Molecular characterization of human rhinovirus field strains isolated during surveillance of enteroviruses. *J Gen Virol* 90, 1371–81.

Bochkov, Y. A., Palmenberg, A. C., Lee, W.-M., Rathe, J. a, Amineva, S. P., Sun, X., Pasic, T. R., Jarjour, N. N., Liggett, S. B., and Gern, J. E. (2011). Molecular modeling, organ culture and reverse genetics for a newly identified human rhinovirus C. *Nat Med* 17, 627–32.

Bouslama, L., Nasri, D., Chollet, L., Belguith, K., Bourlet, T., Aouni, M., Pozzetto, B., and Pillet, S. (2007). Natural recombination event within the capsid genomic region leading to a chimeric strain of human enterovirus B. *J Virol* 81, 8944–52.

Brabec, M., Baravalle, G., Blaas, D., and Fuchs, R. (2003). Conformational changes, plasma membrane penetration, and infection by human rhinovirus type 2: role of receptors and low pH. *J Virol* 77, 5370–7.

Brandenburg, B., Lee, L. Y., Lakadamyali, M., Rust, M. J., Zhuang, X., and Hogle, J. M. (2007). Imaging poliovirus entry in live cells. *PLoS Biol* 5, e183.

Briese, T., Renwick, N., Venter, M., Jarman, R. G., Ghosh, D., Köndgen, S., Shrestha, S. K., Hoegh, A. M., Casas, I., Adjogoua, E. V., et al. (2008). Global distribution of novel rhinovirus genotype. *Emerg Infect Dis* 14, 944–7.

Brittain-Long, R., Westin, J., Olofsson, S., Lindh, M., and Andersson, L.-M. (2010). Prospective evaluation of a novel multiplex real-time PCR assay for detection of fifteen respiratory pathogens-duration of symptoms significantly affects detection rate. *J Clin Virol* 47, 263–7.

Broberg, E., Niemelä, J., Lahti, E., Hyypiä, T., Ruuskanen, O., and Waris, M. (2011). Human rhinovirus C-Associated severe pneumonia in a neonate. *J Clin Virol* 51, 79–82.

Brown, B., Maher, K., Flemister, M. R., Naraghi-Arani, P., Uddin, M., Oberste, M. S., and Pallansch, M. a (2009). Resolving ambiguities in genetic typing of human enterovirus species C clinical isolates and identification of enterovirus 96, 99 and 102. *J Gen Virol* 90, 1713–23.

Brown, B., Oberste, M. S., Maher, K., and Pallansch, M. A. (2003). Complete genomic sequencing shows that polioviruses and members of human enterovirus species C are closely related in the noncapsid coding region. *J Virol* 77, 8973–84.

Brown, D. M., Cornell, C. T., Tran, G. P., Nguyen, J. H. C., and Semler, B. L. (2005). An Authentic 3J Noncoding Region Is Necessary for Efficient Poliovirus Replication. 79, 11962–11973.

Bryce, J., Boschi-Pinto, C., Shibuya, K., and Black, R. E. (2005). WHO estimates of the causes of death in children. *Lancet* 365, 1147–52.

Burns, C. C., Shaw, J., Jorba, J., Bukbuk, D., Adu, F., Gumede, N., Pate, M. A., Abanida, E. A., Gasasira, A., Iber, J., et al. (2013). Multiple Independent Emergences of Type 2 Vaccine-Derived Polioviruses during a Large Outbreak in northern Nigeria. *J Virol*, February [Epub ahead of print].

Calvert, J., Chieochansin, T., Benschop, K. S., McWilliam Leitch, E. C., Drexler, J. F., Grywna, K., da Costa Ribeiro, H., Drosten, C., Harvala, H., Poovorawan, Y., et al. (2010). Recombination dynamics of human parechoviruses: investigation of type-specific differences in frequency and epidemiological correlates. *J Gen Virol* 91, 1229–38.

Chan, Y.-F., Jafar, F. L., Nathan, A.-M., de Bruyne, J. A., Hassan, A., Nor'e, S. S., Abubakar, S., and Sam, I.-C. (2012). Diverse human rhinoviruses A and C from children with respiratory infections in Kuala Lumpur, Malaysia. *J Infect* 64, 633–6.

Chan-Yeung, M., Aït-Khaled, N., White, N., Ip, M. S., and Tan, W. C. (2004). The burden and impact of COPD in Asia and Africa. *Int J Tuberc Lung Dis* 8, 2–14.

Chang, G.-H., Lin, L., Luo, Y.-J., Cai, L.-J., Wu, X.-Y., Xu, H.-M., and Zhu, Q.-Y. (2010). Sequence analysis of six enterovirus 71 strains with different virulences in humans. *Virus Res* 151, 66–73.

Chevaliez, S., Szendröi, A., Caro, V., Balanant, J., Guillot, S., Berencsi, G., and Delpeyroux, F. (2004). Molecular comparison of echovirus 11 strains circulating in Europe during an epidemic of multisystem hemorrhagic disease of infants indicates that evolution generally occurs by recombination. *Virology* 325, 56–70.

Chidlow, G. R., Laing, I. A., Harnett, G. B., Greenhill, A. R., Phuanukoonnon, S., Siba, P. M., Pomat, W. S., Shellam, G. R., Smith, D. W., and Lehmann, D. (2012). Respiratory viral pathogens associated with lower respiratory tract disease among young children in the highlands of Papua New Guinea. *J Clin Virol* 54, 235–239.

Colonno, R. J., Condra, J. H., Mizutani, S., Callahan, P. L., Davies, M. E., and Murcko, M. a (1988). Evidence for the direct involvement of the rhinovirus canyon in receptor binding. *Proc Natl Acad Sci USA* 85, 5449–53.

Cooney, M. K., Fox, J. P., and Kenny, G. E. (1982). Antigenic groupings of 90 rhinovirus serotypes. *Infect Immun* 37, 642–7.

Cordey, S., Gerlach, D., Junier, T., Zdobnov, E. M., Kaiser, L., and Tapparel, C. (2008). The cis-acting replication elements define human enterovirus and rhinovirus species. *RNA* 14, 1568–78.

Costa, C., Bergallo, M., Astegiano, S., Sidoti, F., Terlizzi, M. E., Gambarino, S., Curtoni, A., Simeone, S., Solidoro, P., and Cavallo, R. (2011). Detection of human rhinoviruses in the lower respiratory tract of lung transplant recipients. *Arch Virol* 156, 1439–43.

Crotty, S., Cameron, C. E., and Andino, R. (2001). RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci USA* 98, 6895–900.

Cuervo, N. S., Guillot, S., Romanenkova, N., Combiescu, M., Aubert-Combiescu, A., Seghier, M., Caro, V., Crainic, R., and Delpeyroux, F. (2001). Genomic features of intertypic recombinant sabin poliovirus strains excreted by primary vaccinees. *J Virol* 75, 5740–51.

Dahourou, G., Guillot, S., Le Gall, O., and Crainic, R. (2002). Genetic recombination in wild-type poliovirus. *J Gen Virol* 83, 3103–10.

Dalldorf, G., and Sickles, G. M. (1948). An Unidentified, Filtrable Agent Isolated From the Feces of Children With Paralysis. *Science* 108, 61–2.

Danthi, P., Tosteson, M., Li, Q., and Chow, M. (2003). Genome Delivery and Ion Channel Properties Are Altered in VP4 Mutants of Poliovirus. *J Virol* 77, 5266–5274.

Davis, M. P., Bottley, G., Beales, L. P., Killington, R. a, Rowlands, D. J., and Tuthill, T. J. (2008). Recombinant VP4 of human rhinovirus induces permeability in model membranes. *J Virol* 82, 4169–74.

DeTulleo, L., and Kirchhausen, T. (1998). The clathrin endocytic pathway in viral infection. *EMBO J* 17, 4585–93.

Devaney, M. a, Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E., and Grubman, M. J. (1988). Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *J Virol* 62, 4407–9.

Dierssen, U., Rehren, F., Henke-Gendo, C., Harste, G., and Heim, A. (2008). Rapid routine detection of enterovirus RNA in cerebrospinal fluid by a one-step real-time RT-PCR assay. *J Clin Virol* 42, 58–64.

Domingo, E., Escarmís, C., Sevilla, N., Moya, A., Elena, S. F., Quer, J., Novella, I. S., and Holland, J. J. (1996). Basic concepts in RNA virus evolution. *FASEB J* 10, 859–64.

Domingo, E., Martin, V., Perales, C., and Escarmis, C. (2008). Coxsackieviruses and quasispecies theory: evolution of enteroviruses. *Curr Top Microbiol Immunol* 323, 3–32.

Drake, J. W., and Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci USA* 96, 13910–3.

Dreschers, S., Dumitru, C. a, Adams, C., and Gulbins, E. (2007). The cold case: are rhinoviruses perfectly adapted pathogens? *Cellular and molecular life sciences : CMLS* 64, 181–91.

Drexler, J. F., Baumgarte, S., de Souza Luna, L. K., Stocker, A., Almeida, P. S., Ribeiro, T. C. M., Petersen, N., Herzog, P., Pedroso, C., Brites, C., et al. (2010). Genomic features and evolutionary constraints in Saffold-like cardioviruses. *J Gen Virol* 91, 1418–1427.

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7, 214.

Du, Z., Yu, J., Ulyanov, N. B., Andino, R., and James, T. L. (2004). Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. *Biochemistry* 43, 11959–72.

Duechler, M., Skern, T. I. M., Sommergrubert, W., Neubauer, C., Gruendler, P., Fogyt, I., Blaas, D., and Kuechlert, E. (1987). Evolutionary relationships within the human rhinovirus genus : *Biochemistry* 84, 2605–2609.

Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9, 267–76.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–7.

El-Sahly, H. M., Atmar, R. L., Glezen, W. P., and Greenberg, S. B. (2000). Spectrum of clinical illness in hospitalized patients with "common cold" virus infections. *Clin Infect Dis* 31, 96–100.

Faux, C., Arden, K. E., Lambert, S. B., Nissen, M. D., Nolan, T. M., Chang, A. B., Sloots, T. P., and Mackay, I. M. (2011). Usefulness of published PCR primers in detecting human rhinovirus infection. *Emerg Infect Dis* 17, 296–8.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27, 401–410.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.

Fendrick, a M., Monto, A. S., Nightengale, B., and Sarnes, M. (2003). The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch Intern Med* 163, 487–94.

Fitzgerald, K. D., and Semler, B. L. (2009). Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. *Biochimica et biophysica acta* 1789, 518–28.

Forss, S., and Schaller, H. (1982). A tandem repeat gene in a picornavirus. *Nucleic Acids Res* 10, 6441–50.

Fricks, C. E., and Hogle, J. M. (1990). Cell-induced conformational change in poliovirus: externalization of the amino terminus of VP1 is responsible for liposome binding. *J Virol* 64, 1934–45.

Friedlander, S. L., and Busse, W. W. (2005). The role of rhinovirus in asthma exacerbations. *J Allergy Clin Immunol* 116, 267–73.

Fry, A. M., Lu, X., Olsen, S. J., Chittaganpitch, M., Sawatwong, P., Chantra, S., Baggett, H. C., and Erdman, D. (2011). Human Rhinovirus Infections in Rural Thailand: Epidemiological Evidence for Rhinovirus as Both Pathogen and Bystander. *PloS one* 6, e17780.

Fuji, N., Suzuki, A., Lupisan, S., Sombrero, L., Galang, H., Kamigaki, T., Tamaki, R., Saito, M., Aniceto, R., Olveda, R., et al. (2011). Detection of human rhinovirus C viral genome in blood among children with severe respiratory infections in the Philippines. *PloS one* 6, e27247.

Fujitsuka, A., Tsukagoshi, H., Arakawa, M., Goto-Sugai, K., Ryo, A., Okayama, Y., Mizuta, K., Nishina, A., Yoshizumi, M., Kaburagi, Y., et al. (2011). A molecular epidemiological study of respiratory viruses detected in Japanese children with acute wheezing illness. *BMC Infect Dis* 11, 168.

Gaaloul, I., Riabi, S., Harrath, R., Evans, M., Salem, N. H., Mlayeh, S., Huber, S., and Aouni, M. (2012). Sudden unexpected death related to enterovirus myocarditis: histopathology, immunohistochemistry and molecular pathology diagnosis at post-mortem. *BMC Infect Dis* 12, 212.

Le Gall, O., Christian, P., Fauquet, C. M., King, A. M. Q., Knowles, N. J., Nakashima, N., Stanway, G., and Gorbalenya, A. E. (2008). Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. *Arch Virol* 153, 715–27.

Garbino, J., Soccal, P. M., Aubert, J., Rochat, T., Meylan, P., Thomas, Y., Tapparel, C., Bridevaux, P., and Kaiser, L. (2009). Respiratory viruses in bronchoalveolar lavage : a hospital-based cohort study in adults. *Thorax*, 399–404.

Garriga, D., Pickl-Herk, A., Luque, D., Wruss, J., Castón, J. R., Blaas, D., and Verdaguer, N. (2012). Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. *PLoS pathogens* 8, e1002473.

Gascuel, O., and Steel, M. (2006). Neighbor-joining revealed. *Mol Biol Evol* 23, 1997–2000.

Gaunt, E., McWilliam-Leitch, E. C., Templeton, K., and Simmonds, P. (2009). Incidence, molecular epidemiology and clinical presentations of human metapneumovirus; assessment of its importance as a diagnostic screening target. *J Clin Virol* 46, 318–24.

Gaunt, E. R., Hardie, A., Claas, E. C. J., Simmonds, P., and Templeton, K. E. (2010). Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J Clin Micro* 48, 2940–7.

Gaunt, E. R., Harvala, H., McIntyre, C., Templeton, K. E., and Simmonds, P. (2011). Disease burden of the most commonly detected respiratory viruses in hospitalized patients calculated using the disability adjusted life year (DALY) model. *J Clin Virol* 52, 215–21.

Gerber, K., Wimmer, E., and Paul, A. V. (2001). Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a cis-replicating element in the coding sequence of 2A(pro). *J Virol* 75, 10979–90.

Gern, J. E., and Busse, W. W. (1999). Association of rhinovirus infections with asthma. *Clin Microbiol Rev* 12, 9–18.

Gern, J. E., Galagan, D. M., Jarjour, N. N., Dick, E. C., and Busse, W. W. (1997). Detection of rhinovirus RNA in lower airway cells during experimentally induced infection. *Am J Respir Crit Care Med* 155, 1159–61.

Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–82.

Giranda, V. L., Heinz, B. a, Oliveira, M. a, Minor, I., Kim, K. H., Kolatkar, P. R., Rossmann, M. G., and Rueckert, R. R. (1992). Acid-induced structural changes in human rhinovirus 14: possible role in uncoating. *Proc Natl Acad Sci USA* 89, 10213–7.

Glaser, W., Triendl, A., and Skern, T. (2003). The processing of eIF4GI by human rhinovirus type 2 2A(pro): relationship to self-cleavage and role of zinc. *J Virol* 77, 5021–5.

Gmyl, A. P., Korshenko, S. A., Belousov, E. V., Khitrina, E. V., and Agol, V. I. (2003). Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? *RNA* 9, 1221–31.

Goodfellow, I., Chaudhry, Y., Richardson, A., Meredith, J., Almond, J. W., Barclay, W., and Evans, D. J. (2000). Identification of a cis-acting replication element within the poliovirus coding region. *J Virol* 74, 4590–600.

Greve, J. M., Davis, G., Meyer, a M., Forte, C. P., Yost, S. C., Marlor, C. W., Kamarck, M. E., and McClelland, a (1989). The major human rhinovirus receptor is ICAM-1. *Cell* 56, 839–47.

Guillot, S., Caro, V., Cuervo, N., Korotkova, E., Combiescu, M., Persu, A., Aubert-Combiescu, A., Delpeyroux, F., and Crainic, R. (2000). Natural genetic exchanges between vaccine and wild poliovirus strains in humans. *J Virol* 74, 8434–43.

Guindon, S., and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol* 52, 696–704.

Gutman, J. A., Peck, A. J., Kuypers, J., and Boeckh, M. (2007). Rhinovirus as a cause of fatal lower respiratory tract infection in adult stem cell transplantation patients: a report of two cases. *Bone Marrow Transplant* 40, 809–11.

Hadfield, A., Lee, W. M., Zhao, R., Oliveira, M. a, Minor, I., Rueckert, R. R., and Rossmann, M. G. (1997). The refined structure of human rhinovirus 16 at 2.15 A resolution: implications for the viral life cycle. *Structure* 5, 427–41.

Halfpap, L. M., and Cooney, M. K. (1983). Isolation of rhinovirus intertypes related to either rhinoviruses 12 and 78 or 36 and 58. *Infect Immun* 40, 213–8.

Hamparian, V. V., Colonno, R. J., Cooney, M. K., Dick, E. C., Gwaltney, J. M., Hughes, J. H., Jordan, W. S., Kapikian, A. Z., Mogabgab, W. J., and Monto, A. (1987). A collaborative report: rhinoviruses--extension of the numbering system from 89 to 100. *Virology* 159, 191–2.

Han, T. H., Chung, J. Y., Hwang, E. S., and Koo, J. W. (2009). Detection of human rhinovirus C in children with acute lower respiratory tract infections in South Korea. *Arch Virol* 154, 987–91.

Harris, J. R., and Racaniello, V. R. (2005). Amino acid changes in proteins 2B and 3A mediate rhinovirus type 39 growth in mouse cells. *J Virol* 79, 5363–73.

Harvala, H., McIntyre, C. L., Imai, N., Clasper, L., Djoko, C. F., LeBreton, M., Vermeulen, M., Saville, A., Mutapi, F., Tamoufé, U., et al. (2012a). High seroprevalence of enterovirus infections in apes and old world monkeys. *Emerg Infect Dis* 18, 283–6.

Harvala, H., McIntyre, C. L., McLeish, N. J., Kondracka, J., Palmer, J., Molyneaux, P., Gunson, R., Bennett, S., Templeton, K., and Simmonds, P. (2012b). High detection frequency and viral loads of human rhinovirus species A to C in fecal samples; diagnostic and clinical implications. *J Med Virol* 84, 536–42.

Harvala, H., Sharp, C. P., Ngole, E. M., Delaporte, E., Peeters, M., and Simmonds, P. (2011). Detection and genetic characterization of enteroviruses circulating among wild

populations of chimpanzees in Cameroon: relationship with human and simian enteroviruses. *J Virol* 85, 4480–6.

Heath, L., van der Walt, E., Varsani, A., and Martin, D. P. (2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* 80, 11827–32.

Hellen, C. U. T., and de Breyne, S. (2007). A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse picornavirus genera: evidence for modular exchange of functional noncoding RNA elements by recombination. *J Virol* 81, 5850–63.

Henquell, C., Mirand, A., Deusebis, A.-L., Regagnon, C., Archimbaud, C., Chambon, M., Bailly, J.-L., Gourdon, F., Hermet, E., Dauphin, J.-B., et al. (2012). Prospective genotyping of human rhinoviruses in children and adults during the winter of 2009-2010. *J Clin Virol* 53, 280–4.

Herold, J., and Andino, R. (2001). Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol Cell* 7, 581–91.

Hewat, E. A., Neumann, E., Conway, J. F., Moser, R., Ronacher, B., Marlovits, T. C., and Blaas, D. (2000). The cellular receptor to human rhinovirus 2 binds around the 5-fold axis and not in the canyon : a structural view. *EMBO J* 19, 6317–6325.

Hicks, A. L., and Duffy, S. (2011). Genus-Specific Substitution Rate Variability among Picornaviruses. *J Virol* 85, 7942–7.

Hicks, L. A., Shepard, C. W., Britz, P. H., Erdman, D. D., Fischer, M., Flannery, B. L., Peck, A. J., Lu, X., Thacker, W. L., Benson, R. F., et al. (2006). Two outbreaks of severe respiratory disease in nursing homes associated with rhinovirus. *J Am Ger Soc* 54, 284–9.

Hofer, F., Gruenberger, M., Kowalski, H., Machat, H., Huettinger, M., Kuechler, E., and Blaas, D. (1994). Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc Natl Acad Sci USA* 91, 1839–42.

Holder, M., and Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* 4, 275–84.

Holland, D. T., Senne, J., Peter, C. R., Urmenta, C., and Connor, J. D. (1998). Differentiation and characterization of enteroviruses by computer-assisted viral protein fingerprinting. *J Clin Micro* 36, 1588–94.

Honkanen, H., Oikarinen, S., Peltonen, P., Simell, O., Ilonen, J., Veijola, R., Knip, M., and Hyoty, H. (2013). Human rhinoviruses including group C are common in stool samples of young Finnish children. *J Clin Virol* 56, 334–338.

Horsnell, C., Gama, R. E., Hughes, P. J., and Stanway, G. (1995). Molecular relationships between 21 human rhinovirus serotypes. *J Gen Virol* 76, 2549–55.

Huang, S.-C., Hsu, Y.-W., Wang, H.-C., Huang, S.-W., Kiang, D., Tsai, H.-P., Wang, S.-M., Liu, C.-C., Lin, K.-H., Su, I.-J., et al. (2008). Appearance of intratypic recombination of enterovirus 71 in Taiwan from 2002 to 2005. *Virus Res* 131, 250–9.

Huang, T., Wang, W., Bessaud, M., Ren, P., Sheng, J., Yan, H., Zhang, J., Lin, X., Wang, Y., Delpeyroux, F., et al. (2009). Evidence of recombination and genetic diversity in human rhinoviruses in children with acute respiratory infection. *PLoS One* 4, e6355.

Hughes, P. J., North, C., Jellis, C. H., Minor, P. D., and Stanway, G. (1988). The nucleotide sequence of human rhinovirus 1B: molecular relationships within the rhinovirus genus. *J Gen Virol* 69 (Pt 1), 49–58.

Hyypiä, T., Hovi, T., Knowles, N. J., and Stanway, G. (1997). Classification of enteroviruses based on molecular and biological properties. *J Gen Virol* 78 (Pt 1), 1–11.

Ishiko, H., Miura, R., Shimada, Y., Hayashi, A., Nakajima, H., Yamazaki, S., and Takeda, N. (2002). Human rhinovirus 87 identified as human enterovirus 68 by VP4-based molecular diagnosis. *Intervirology* 45, 136–41.

JG Kapsenberg, A Ras, J. K. (1980). Improvement of enterovirus neutralization by treatment with sodium deoxycholate or chloroform. *Intervirology* 12, 329–34.

Jackson, D. J., Gangnon, R. E., Evans, M. D., Roberg, K. a, Anderson, E. L., Pappas, T. E., Printz, M. C., Lee, W.-M., Shult, P. a, Reisdorf, E., et al. (2008). Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. *Am J Respir Crit Care Med* 178, 667–72.

Jang, S. K., Kräusslich, H. G., Nicklin, M. J., Duke, G. M., Palmenberg, a C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* 62, 2636–43.

Jartti, T., and Korppi, M. (2011). Rhinovirus-induced bronchiolitis and asthma development. *Paediatr Allergy Immunol* 22, 350–5.

Jegouic, S., Joffret, M.-L., Blanchard, C., Riquet, F. B., Perret, C., Pelletier, I., Colbere-Garapin, F., Rakoto-Andrianarivelo, M., and Delpeyroux, F. (2009). Recombination between polioviruses and co-circulating Coxsackie A viruses: role in the emergence of pathogenic vaccine-derived polioviruses. *PLoS Pathog* 5, e1000412.

Jenkins, G. M., Rambaut, A., Pybus, O. G., and Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 54, 156–65.

Jin, Y., Yuan, X.-H., Xie, Z.-P., Gao, H.-C., Song, J.-R., Zhang, R.-F., Xu, Z.-Q., Zheng, L.-S., Hou, Y.-D., and Duan, Z.-J. (2009). Prevalence and clinical characterization of a newly identified human rhinovirus C species in children with acute respiratory tract infections. *J Clin Micro* 47, 2895–900.

Joshi, M. S., Walimbe, A. M., and Chitambar, S. D. (2008). Evaluation of genomic regions of hepatitis A virus for phylogenetic analysis: Suitability of the 2C region for genotyping. *J Virol Methods* 153, 36–42.

Kaida, A., Kubo, H., Takakura, K., Togawa, M., Shiomi, M., Kohdera, U., and Iritani, N. (2011). Molecular epidemiology of human rhinovirus C in patients with acute respiratory tract infections in Osaka City, Japan. *Jpn J Infect Dis* 64, 488–92.

Kapikian, A., Conant, R., and Hamparian, V. (1967). Rhinoviruses: a numbering system. *Nature* 213, 761–2.

Kapikian, A., Conant, R., Hamparian, V., Chanock, R., Dick, E., Gwaltney, J., Hamre, D., Jordan, W., Kenny, G., Lennette, E., et al. (1971). A collaborative report: rhinoviruses--extension of the numbering system. *Virology* 43, 524–526.

Katpally, U., Fu, T.-M., Freed, D. C., Casimiro, D. R., and Smith, T. J. (2009). Antibodies to the buried N terminus of rhinovirus VP4 exhibit cross-serotypic neutralization. *J Virol* 83, 7040–8.

Khetsuriani, N., Lu, X., Teague, W. G., Kazerouni, N., Anderson, L. J., and Erdman, D. D. (2008). Novel human rhinoviruses and exacerbation of asthma in children. *Emerg Infect Dis* 14, 1793–6.

Kiang, D., Kalra, I., Yagi, S., Louie, J. K., Boushey, H., Boothby, J., and Schnurr, D. P. (2008). Assay for 5' noncoding region analysis of all human rhinovirus prototype strains. *J Clin Microbiol* 46, 3736–45.

Kiang, D., Newbower, E. C., Yeh, E., Wold, L., Chen, L., and Schnurr, D. P. (2009). An algorithm for the typing of enteroviruses and correlation to serotyping by viral neutralization. *J Clin Virol* 45, 334–40.

Kilpatrick, D. R., Nottay, B., Yang, C. F., Yang, S. J., Da Silva, E., Peñaranda, S., Pallansch, M., and Kew, O. (1998). Serotype-specific identification of polioviruses by PCR using primers containing mixed-base or deoxyinosine residues at positions of codon degeneracy. *J Clin Micro* 36, 352–7.

Kirkegaard, K., and Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. *Cell* 47, 433–443.

Kistler, A., Avila, P. C., Rouskin, S., Wang, D., Ward, T., Yagi, S., Schnurr, D., Ganem, D., DeRisi, J. L., and Boushey, H. A. (2007a). Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* 196, 817–25.

Kistler, A. L., Webster, D. R., Rouskin, S., Magrini, V., Credle, J. J., Schnurr, D. P., Boushey, H. A., Mardis, E. R., Li, H., and DeRisi, J. L. (2007b). Genome-wide diversity and selective pressure in the human rhinovirus. *Virol J* 4, 40.

Knowles, N. J., Hovi, T., Hyypiä, T., King, A. M. Q., Lindberg, A. M., Pallansch, M. A., Palmenberg, A. C., Simmonds, P., Skern, T., Stanway, G., et al. (2012). *Picornaviridae. In Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. , eds. A. M. Q. King, M. J. Adams, E. B. Carstens, and E. J. Lefkowitz. Elsevier.

Knowles, N., and Wadsworth, J. (2010). The complete genome sequence of a picornavirus isolated from a harbour (common) seal (Phoca vitulina). in *EUROPIC : XVI Meeting of the European Study Group on the Molecular Biology of Picornaviruses, St. Andrews, Scotland*.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. (2006a). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23, 1891–901.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. (2006b). GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22, 3096–8.

Kroneman, A., Vennema, H., Deforche, K., v d Avoort, H., Peñaranda, S., Oberste, M. S., Vinjé, J., and Koopmans, M. (2011). An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 51, 121–5.

Kupila, L., Vuorinen, T., Vainionpää, R., Hukkanen, V., Marttila, R. J., and Kotilainen, P. (2006). Etiology of aseptic meningitis and encephalitis in an adult population. *Neurology* 66, 75–80.

Kusel, M. M., de Klerk, N. H., Holt, P. G., Kebadze, T., Johnston, S. L., and Sly, P. D. (2006). Role of respiratory viruses in acute upper and lower respiratory tract illness in the first year of life: a birth cohort study. *Pediatr Infect Dis J* 25, 680–6.

Kwok, S., and Higuchi, R. (1989). Avoiding false positives with PCR. *Nature* 339, 237–8.

Lai, M. M. (1992). RNA recombination in animal and plant viruses. *Microbiol Rev* 56, 61–79.

Laine, P., Blomqvist, S., Savolainen, C., Andries, K., and Hovi, T. (2006). Alignment of capsid protein VP1 sequences of all human rhinovirus prototype strains: conserved motifs and functional domains. *J Gen Virol* 87, 129–38.

Laine, P., Savolainen, C., Blomqvist, S., and Hovi, T. (2005). Phylogenetic analysis of human rhinovirus capsid protein VP1 and 2A protease coding sequences confirms shared genus-like relationships with human enteroviruses. *J Gen Virol* 86, 697–706.

Lamson, D., Renwick, N., Kapoor, V., Liu, Z., Palacios, G., Ju, J., Dean, A., St George, K., Briese, T., and Lipkin, W. I. (2006). MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004-2005. *J Infect Dis* 194, 1398–402.

Lau, S. K. P., Yip, C. C. Y., Lung, D. C., Lee, P., Que, T.-L., Lau, Y.-L., Chan, K.-H., Woo, P. C. Y., and Yuen, K.-Y. (2012). Detection of human rhinovirus C in fecal samples of children with gastroenteritis. *J Clin Virol* 53, 290–6.

Lau, S. K. P., Yip, C. C. Y., Tsoi, H.-W., Lee, R. a, So, L.-Y., Lau, Y.-L., Chan, K.-H., Woo, P. C. Y., and Yuen, K.-Y. (2007). Clinical features and complete genome characterization of a distinct human rhinovirus (HRV) genetic cluster, probably representing a previously undetected HRV species, HRV-C, associated with acute respiratory illness in children. *J Clin Micro* 45, 3655–64.

Ledford, R. M., Collett, M. S., and Pevear, D. C. (2005). Insights into the genetic basis for natural phenotypic resistance of human rhinoviruses to pleconaril. *Antiviral Res* 68, 135–8.

Ledford, R. M., Patel, N. R., Demenczuk, T. M., Watanyar, A., Herbertz, T., Collett, M. S., Pevear, D. C., and Irol, J. V. (2004). VP1 Sequencing of All Human Rhinovirus Serotypes : Insights into Genus Phylogeny and Susceptibility to Antiviral Capsid-Binding Compounds. *J Virol* 78, 3663–3674.

Lee, W. M., Kiesner, C., Pappas, T., Lee, I., Grindle, K., Jartti, T., Jakiela, B., Lemanske, R. F., Shult, P. A., and Gern, J. E. (2007). A diverse group of previously unrecognized human rhinoviruses are common causes of respiratory illnesses in infants. *PLoS One* 3, e966.

Lee, W. M., Monroe, S. S., and Rueckert, R. R. (1993). Role of maturation cleavage in infectivity of picornaviruses: activation of an infectosome. *J Virol* 67, 2110–22.

Lee, W. M., Wang, W., and Rueckert, R. R. (1995). Complete sequence of the RNA genome of human rhinovirus 16, a clinically useful common cold virus belonging to the ICAM-1 receptor group. *Virus genes* 9, 177–81.

Lewis, J. K., Bothner, B., Smith, T. J., and Siuzdak, G. (1998). Antiviral agent blocks breathing of the common cold virus. *Proc Natl Acad Sci USA* 95, 6774–8.

Lewis-Rogers, N., Bendall, M. L., and Crandall, K. A. (2009). Phylogenetic relationships and molecular adaptation dynamics of human rhinoviruses. *Mol Biol Evol* 26, 969–81.

Li, Q., Yafal, a G., Lee, Y. M., Hogle, J., and Chow, M. (1994). Poliovirus neutralization by antibodies to internal epitopes of VP4 and VP1 results from reversible exposure of these sequences at physiological temperature. *J Virol* 68, 3965–70.

Lindberg, A. M., Andersson, P., Savolainen, C., Mulders, M. N., and Hovi, T. (2003). Evolution of the genome of Human enterovirus B: incongruence between phylogenies of the VP1 and 3CD regions indicates frequent recombination within the species. *J Gen Virol* 84, 1223–1235.

Linsuwanon, P., Payungporn, S., Samransamruajkit, R., Posuwan, N., Makkoch, J., Theanboonlers, A., and Poovorawan, Y. (2009). High prevalence of human rhinovirus C infection in Thai children with acute lower respiratory tract disease. *J Infect* 59, 115–21.

Linsuwanon, P., Payungporn, S., Suwannakarn, K., Chieochansin, T., Theamboonlers, A., and Poovorawan, Y. (2011). Complete coding sequence characterization and comparative analysis of the putative novel human rhinovirus (HRV) species C and B. *Virol J* 8, 5.

Lloyd, R. E., Grubman, M. J., and Ehrenfeld, E. (1988). Relationship of p220 cleavage during picornavirus infection to 2A proteinase sequencing. *J Virol* 62, 4216–23.

van Loon, A., Cleator, G., and Ras, A. (1999). External quality assessment of enterovirus detection and typing. European Union Concerted Action of Virus Meningitis and Encephalitis. *Bull World Health Organ* 77, 217–223.

Lukashev, A. N. (2005). Role of recombination in evolution of enteroviruses. *Rev Med Virol* 15, 157–67.

Lukashev, A. N., Lashkevich, V. A., Ivanova, O. E., Koroleva, G. A., Hinkkanen, A. E., and Ilonen, J. (2005). Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. *J Gen Virol* 86, 3281–90.

Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., Lindberg, a M., Persson, B., Allander, T., and Andersson, B. (2012). Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PloS one* 7, e30875.

Macadam, A., Arnold, C., Howlett, J., John, A., Marsden, S., Taffs, F., Reeve, P., Hamada, N., Wareham, K., and Almond, J. (1989). Reversion of the attenuated and temperature-sensitive phenotypes of the Sabin type 3 strain of poliovirus in vaccinees. *Virology* 172, 408–14.

Mallia, P., Message, S. D., Gielen, V., Contoli, M., Gray, K., Kebadze, T., Aniscenko, J., Laza-Stanca, V., Edwards, M. R., Slater, L., et al. (2011). Experimental rhinovirus infection as a human model of chronic obstructive pulmonary disease exacerbation. *Am J Respir Crit Care Med* 183, 734–42.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–3.

Martin, D. P., Posada, D., Crandall, K. A., and Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* 21, 98–102.

Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–3.

McErlean, P., Shackelton, L. a, Andrews, E., Webster, D. R., Lambert, S. B., Nissen, M. D., Sloots, T. P., and Mackay, I. M. (2008). Distinguishing molecular features and clinical characteristics of a putative new rhinovirus species, human rhinovirus C (HRV C). *PloS one* 3, e1847.

McErlean, P., Shackelton, L. a, Lambert, S. B., Nissen, M. D., Sloots, T. P., and Mackay, I. M. (2007). Characterisation of a newly identified human rhinovirus, HRV-QPM, discovered in infants with bronchiolitis. *J Clin Virol* 39, 67–75.

McIntyre, C., Knowles, N., and Simmonds, P. (2013a). Proposals for the classification of human rhinovirus species A, B and C (HRV-A, -B and –C) into genotypically assigned types. [Submitted].

McIntyre, C., McWilliam Leitch, E. C., Savolainen-Kopra, C., Hovi, T., and Simmonds, P. (2010). Analysis of genetic diversity and sites of recombination in human rhinovirus species C. *J Virol* 84, 10297–310.

McIntyre, C., Savolainen-Kopra, C., Hovi, T., and Simmonds, P. (2013b). Recombination in the evolution of human rhinovirus genomes. *Arch Virol*, February [Epub ahead of print].

McKnight, K. L., and Lemon, S. M. (1998). The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA* 4, 1569–84.

McLeish, N., Witteveldt, J., Clasper, L., McIntyre, C., McWilliam Leitch, E. C., Hardie, A., Bennett, S., Gunson, R., Carman, W. F., Feeney, S. a, et al. (2012). Development and assay of RNA transcripts of enterovirus species A to D, rhinovirus species a to C, and human parechovirus: assessment of assay sensitivity and specificity of real-time screening and typing methods. *J Clin Micro* 50, 2910–7.

McManus, T. E., Marley, A.-M., Baxter, N., Christie, S. N., O'Neill, H. J., Elborn, J. S., Coyle, P. V., and Kidney, J. C. (2008). Respiratory viral infection in exacerbations of COPD. *Respir Med* 102, 1575–80.

McWilliam Leitch, E. C., Bendig, J., Cabrerizo, M., Cardosa, J., Hyypiä, T., Ivanova, O. E., Kelly, A., Kroes, A. C., Lukashev, A., MacAdam, A., et al. (2009a). Transmission networks and population turnover of echovirus 30. *J Virol* 83, 2109–18.

McWilliam Leitch, E. C., Cabrerizo, M., Cardosa, J., Harvala, H., Ivanova, O. E., Koike, S., Kroes, A. C., Lukashev, A. N., Perera, D., Roivainen, M., et al. (2012). The association of recombination events in the founding and emergence of subgenogroup evolutionary lineages of human enterovirus 71. *J Virol* 86, 2676–85.

McWilliam Leitch, E. C., Cabrerizo, M., Cardosa, J., Harvala, H., Ivanova, O. E., Kroes, A. C., Lukashev, A., Muir, P., Odoom, J., Roivainen, M., et al. (2010). Evolutionary dynamics and temporal/geographical correlates of recombination in the human enterovirus echovirus types 9, 11, and 30. *J Virol* 84, 9292–300.

McWilliam Leitch, E. C., Harvala, H., Robertson, I., Ubillos, I., Templeton, K., and Simmonds, P. (2009b). Direct identification of human enterovirus serotypes in cerebrospinal fluid by amplification and sequencing of the VP1 region. *J Clin Microbiol* 44, 119–124.

Medappa, K. C., McLean, C., and Rueckert, R. R. (1971). On the structure of rhinovirus 1A. *Virology* 44, 259–70.

Melnick, J. L., Shaw, E. W., and Curnen, E. C. (1949). A virus isolated from patients diagnosed as non-paralytic poliomyelitis or aseptic meningitis. *Proc Soc Exp Biol Med* 71, 344–9.

Miller, E. K., Edwards, K. M., Weinberg, G. A., Iwane, M. K., Griffin, M. R., Hall, C. B., Zhu, Y., Szilagyi, P. G., Morin, L. L., Heil, L. H., et al. (2009a). A novel group of rhinoviruses is associated with asthma hospitalizations. *J Allergy Clin Immunol* 123, 98–104.e1.

Miller, E. K., Khuri-Bulos, N., Williams, J. V., Shehabi, A. a, Faouri, S., Al Jundi, I., Chen, Q., Heil, L., Mohamed, Y., Morin, L.-L., et al. (2009b). Human rhinovirus C associated with wheezing in hospitalised children in the Middle East. *J Clin Virol* 46, 85–9.

Mirand, A., Rouveyrol, F., Chambon, M., Archimbaud, C., Regagnon, C., Bailly, J.-L., Peigue-Lafeuille, H., and Henquell, C. (2009). Enterovirus genotyping directly from original clinical specimens: prospective application to a severe neonatal infection. *J Clin Virol* 44, 177–8.

Mizuta, K., Hirata, A., Suto, A., Aoki, Y., Ahiko, T., Itagaki, T., Tsukagoshi, H., Morita, Y., Obuchi, M., Akiyama, M., et al. (2010). Phylogenetic and cluster analysis of human rhinovirus species A (HRV-A) isolated from children with acute respiratory infections in Yamagata, Japan. *Virus Res* 147, 265–74.

Mosser, A. G., Brockman-Schneider, R., Amineva, S., Burchell, L., Sedgwick, J. B., Busse, W. W., and Gern, J. E. (2002). Similar frequency of rhinovirus-infectible cells in upper and lower airway epithelium. *The Journal of infectious diseases* 185, 734–43.

Nix, W. A., Oberste, M. S., and Pallansch, M. a (2006). Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. *J Clin Micro* 44, 2698–704.

Nokso-Koivisto, J., Hovi, T., and Pitkäranta, A. (2006). Viral upper respiratory tract infections in young children with emphasis on acute otitis media. *Int J Paediatr Otorhinolaryngol* 70, 1333–42.

Norder, H., Bjerregaard, L., and Magnius, L. O. (2002). Open reading frame sequence of an Asian enterovirus 73 strain reveals that the prototype from California is recombinant. *J Gen Virol* 83, 1721–8.

Nyquist, A., Gonzales, R., Steiner, J. F., and Sande, M. a (1998). Antibiotic prescribing for children with colds, upper respiratory tract infections, and bronchitis. *JAMA* 279, 875–7.

Oberste, M. S., Maher, K., Flemister, M. R., Marchetti, G., Kilpatrick, D. R., and Pallansch, M. a (2000). Comparison of classic and molecular approaches for the identification of untypeable enteroviruses. *J Clin Micro* 38, 1170–4.

Oberste, M. S., Maher, K., Kilpatrick, D. R., Flemister, M. R., Brown, B. A., and Pallansch, M. A. (1999a). Typing of human enteroviruses by partial sequencing of VP1. *J Clin Microbiol* 37, 1288–93.

Oberste, M. S., Maher, K., Kilpatrick, D. R., and Pallansch, M. A. (1999b). Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J Virol* 73, 1941–8.

Oberste, M. S., Maher, K., Nix, W. a, Michele, S. M., Uddin, M., Schnurr, D., Al-Busaidy, S., Akoua-Koffi, C., and Pallansch, M. a (2007). Molecular identification of 13 new

enterovirus types, EV79-88, EV97, and EV100-101, members of the species Human Enterovirus B. *Virus Res* 128, 34–42.

Oberste, M. S., Maher, K., and Pallansch, M. A. (2004a). Evidence for frequent recombination within species human enterovirus B based on complete genomic sequences of all thirty-seven serotypes. *J Virol* 78, 855–67.

Oberste, M. S., Maher, K., and Pallansch, M. A. (2002). Molecular Phylogeny and Proposed Classification of the Simian Picornaviruses. *J Virol* 76, 1244–1251.

Oberste, M. S., Maher, K., and Pallansch, M. a (1998). Molecular phylogeny of all human enterovirus serotypes based on comparison of sequences at the 5' end of the region encoding VP2. *Virus Res* 58, 35–43.

Oberste, M. S., Maher, K., Schnurr, D., Flemister, M. R., Lovchik, J. C., Peters, H., Sessions, W., Kirk, C., Chatterjee, N., Fuller, S., et al. (2004b). Enterovirus 68 is associated with respiratory illness and shares biological features with both the enteroviruses and the rhinoviruses. *J Gen Virol* 85, 2577–84.

Oberste, M. S., Michele, S. M., Maher, K., Schnurr, D., Cisterna, D., Junttila, N., Uddin, M., Chomel, J.-J., Lau, C.-S., Ridha, W., et al. (2004c). Molecular identification and characterization of two proposed new enterovirus serotypes, EV74 and EV75. *J Gen Virol* 85, 3205–12.

Oberste, M. S., Peñaranda, S., and Pallansch, M. A. (2004d). RNA recombination plays a major role in genomic change during circulation of coxsackie B viruses. *J Virol* 78, 2948–55.

Ohka, S., Yang, W., Terada, E., Iwasaki, K., and Nomoto, A. (1998). Retrograde transport of intact poliovirus through the axon via the fast transport system. *Virology* 1, 67–75.

Oliveira, M., Zhao, R., Lee, W., Kremer, M., Minor, I., Rueckert, R., Diana, G., Pevear, D., Dutko, F., and Mckinlay, M. (1993). The structure of human rhinovirus 16. *Structure* 1, 51–68.

Oprisan, G., Combiescu, M., Guillot, S., Caro, V., Combiescu, A., Delpeyroux, F., and Crainic, R. (2002). Natural genetic recombination between co-circulating heterotypic enteroviruses. *J Gen Virol* 83, 2193–200.

Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218–25.

Pallansch, M., and Roos, R. (2007). "Enteroviruses: polioviruses, coxsackieviruses, echoviruses and newer enteroviruses.," in *Field Virology, 5th ed.*, eds. D. Knipe and P. Howley (Philadelphia: Lippincott Williams & Wilkins), 839–93.

Palmenberg, A., Rathe, J. A., and Liggett, S. B. (2010). Analysis of the complete genome sequences of human rhinovirus. *J Allergy Clin Immunol* 125, 1190–9.

Palmenberg, A., Spiro, D., Kuzmickas, R., Wang, S., Djikeng, A., Rathe, J. A., Fraser-Liggett, C. M., and Liggett, S. B. (2009). Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. *Science* 324, 55–9.

Papadopoulos, N. G. (2002). Association of Rhinovirus Infection with Increased Disease Severity in Acute Bronchiolitis. *Am J Respir Crit Care Med* 165, 1285–1289.

Papadopoulos, N. G., Bates, P. J., Bardin, P. G., Papi, A., Leir, S. H., Fraenkel, D. J., Meyer, J., Lackie, P. M., Sanderson, G., Holgate, S. T., et al. (2000). Rhinoviruses infect the lower airways. *J Infect Dis* 181, 1875–84.

Papadopoulos, N. G., Sanderson, G., Hunter, J., and Johnston, S. L. (1999). Rhinoviruses replicate effectively at lower airway temperatures. *J Med Virol* 58, 100–4.

Parody, R., Rabella, N., Martino, R., Otegui, M., del Cuerpo, M., Coll, P., and Sierra, J. (2007). Upper and lower respiratory tract infections by human enterovirus and rhinovirus in adult patients with hematological malignancies. *Am J Hematol* 82, 807–11.

Paul, A., Rieder, E., Kim, D. W., van Boom, J. H., and Wimmer, E. (2000). Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *J Virol* 74, 10359–70.

Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320–5.

Pelon, W., Mogabgab, W. J., Phillip, I. A., and Pierce, W. E. (1957). A cytopathogenic agent isolated from naval recruits with mild respiratory illnesses. *Proc Soc Exp Biol Med* 94, 262–7.

Peltola, V., Waris, M., Osterback, R., Susi, P., Ruuskanen, O., and Hyypiä, T. (2008). Rhinovirus transmission within families with children: incidence of symptomatic and asymptomatic infections. *J Infect Dis* 197, 382–9.

Petersen, J. F., Cherney, M. M., Liebig, H. D., Skern, T., Kuechler, E., and James, M. N. (1999). The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J* 18, 5463–75.

Pfeiffer, J. K., and Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* 1, e11.

Piralla, A., Baldanti, F., and Gerna, G. (2011). Phylogenetic patterns of human respiratory picornavirus species, including the newly identified group C rhinoviruses, during a 1-year surveillance of a hospitalized patient population in Italy. *J Clin Micro* 49, 373–6.

Piralla, A., Rovida, F., Campanini, G., Rognoni, V., Marchi, A., Locatelli, F., and Gerna, G. (2009). Clinical severity and molecular typing of human rhinovirus C strains during a fall outbreak affecting hospitalized patients. *J Clin Virol* 45, 311–7.

Pitkäranta, a, Starck, M., Savolainen, S., Pöyry, T., Suomalainen, I., Hyypiä, T., Carpen, O., and Vaheri, A. (2001). Rhinovirus RNA in the maxillary sinus epithelium of adult patients with acute sinusitis. *Clin Infect Dis* 33, 909–11.

Pitkäranta, a, Virolainen, a, Jero, J., Arruda, E., and Hayden, F. G. (1998). Detection of rhinovirus, respiratory syncytial virus, and coronavirus infections in acute otitis media by reverse transcriptase polymerase chain reaction. *Pediatrics* 102, 291–5.

Pitkäranta, A., Arruda, E., Malmberg, H., Hayden, F. G., Arruda, E., Malmberg, H., and Pitka, A. (1997). Detection of rhinovirus in sinus brushings of patients with acute community-acquired sinusitis by reverse transcription-PCR . These include : Detection of Rhinovirus in Sinus Brushings of Patients with Acute Community-Acquired Sinusitis by Reverse Transcr. *Microbiology* 35, 7–10.

Posada, D. (2002). Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* 19, 708–17.

Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 98, 13757–62.

du Prel, J.-B., Puppe, W., Gröndahl, B., Knuf, M., Weigl, J. a I., Schaaff, F., and Schmitt, H.-J. (2009). Are meteorological parameters associated with acute respiratory tract infections? *Clin Infect Dis* 49, 861–8.

Price, W. H. (1956). The isolation of a new virus associated with respiratory clinical disease in humans. *Proc Natl Acad Sci USA* 42, 892–6.

Pöyry, T., Hyypiä, T., Horsnell, C., Kinnunen, L., Hovi, T., and Stanway, G. (1994). Molecular analysis of coxsackievirus A16 reveals a new genetic group of enteroviruses. *Virology* 202, 982–7.

Pöyry, T., Kinnunen, L., Hyypiä, T., Brown, B., Horsnell, C., Hovi, T., and Stanway, G. (1996). Genetic and phylogenetic clustering of enteroviruses. *J Gen Virol* 77 (Pt 8), 1699–717.

VR Racaniello (2007). "Picornaviridae : the viruses and their replication," in *Fields Virology*, eds. D. Knipe and P. Howley (Philadelphia: Lippincott Williams & Wilkins), 795 – 834.

Rathe, J. A., Liu, X., Tallon, L. J., Gern, J. E., and Liggett, S. B. (2010). Full-genome sequence and analysis of a novel human rhinovirus strain within a divergent HRV-A clade. *Arch Virol* 155, 83–7.

Ren, R., and Racaniello, V. (1992). Poliovirus spreads from muscle to the central nervous system by neural pathways. *J Infect Dis* 4, 747–52.

Renwick, N., Schweiger, B., Kapoor, V., Liu, Z., Villari, J., Bullmann, R., Miething, R., Briese, T., and Lipkin, W. I. (2007). A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. *J Infect Dis* 196, 1754–60.

Robbins, F. C., Enders, J. F., Weller, T. H., and Florentino, G. L. (1951). Studies on the cultivation of poliomyelitis viruses in tissue culture. V. The direct isolation and serologic identification of virus strains in tissue culture from patients with nonparalytic and paralytic poliomyelitis. *Am J Hyg* 54, 286–93.

Rohll, J. B., Percy, N., Ley, R., Evans, D. J., Almond, J. W., and Barclay, W. S. (1994). The 5'-untranslated regions of picornavirus RNAs contain independent functional domains essential for RNA replication and translation. *J Virol* 68, 4384–91.

Roy, A., and Post, C. B. (2012). Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity. *Proc Natl Acad Sci USA* 109, 5271–6.

Rueckert, R. R., and Wimmer, E. (1984). Systematic nomenclature of picornavirus proteins. *J Virol* 50, 957–9.

Runckel, C., Westesson, O., Andino, R., and Derisi, J. L. (2013). Identification and manipulation of the molecular determinants influencing poliovirus recombination. *PLoS pathogens* 9, e1003164.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–25.

van der Sanden, S., van Eek, J., Martin, D. P., van der Avoort, H., Vennema, H., and Koopmans, M. (2011). Detection of recombination breakpoints in the genomes of human enterovirus 71 strains isolated in the Netherlands in epidemic and non-epidemic years, 1963-2010. *Infect Genet Evol* 11, 886–94.

Sanfaçon, H., Gorbalenya, A. E., Knowles, N. J., and Chen, Y. (2011). *Picornavirales. In Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the*

International Committee on Taxonomy of Viruses. , eds. A. M. Q. King, M. J. Adams, E. B. Carstens, and E. . Lefkowitz San Diego: Elsevier.

Santti, J., Hyypiä, T., Kinnunen, L., and Salminen, M. (1999). Evidence of recombination among enteroviruses. *J Virol* 73, 8741–9.

Savolainen, C., Laine, P., Mulders, M. N., and Hovi, T. (2004). Sequence analysis of human rhinoviruses in the RNA-dependent RNA polymerase coding region reveals large within-species variation. *J Gen Virol* 85, 2271–7.

Savolainen, C., Mulders, M. N., and Hovi, T. (2002). Phylogenetic analysis of rhinovirus isolates collected during successive epidemic seasons. *Virus Res* 85, 41–6.

Savolainen-Kopra, C., Blomqvist, S., Kilpi, T., Roivainen, M., and Hovi, T. (2009a). Novel species of human rhinoviruses in acute otitis media. *Pediatr Infect Dis J* 28, 59–61.

Savolainen-Kopra, C., Blomqvist, S., Smura, T., Roivainen, M., Hovi, T., Kiang, D., Kalra, I., Yagi, S., Louie, J. K., Boushey, H., et al. (2009b). 5' Noncoding Region Alone Does Not Unequivocally Determine Genetic Type of Human Rhinovirus Strains. *J Clin Micro* 47, 1278–80.

Scheltinga, S. A., Templeton, K. E., Beersma, M. F. C., and Claas, E. C. J. (2005). Diagnosis of human metapneumovirus and rhinovirus in patients with respiratory tract infections by an internally controlled multiplex real-time RNA PCR. *J Clin Virol* 33, 306–11.

Schibler, M., Gerlach, D., Martinez, Y., Belle, S. V., Turin, L., Kaiser, L., and Tapparel, C. (2012). Experimental human rhinovirus and enterovirus interspecies recombination. *J Gen Virol* 93, 93–101.

Schmidt, H. a, Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–4.

Schmidt, N. J., Lennette, E. H., and Ho, H. H. (1966). Observations on antigenic variants of echovirus type 11. *Proc Soc Exp Biol Med* 123, 696–700.

Shapiro, B., Rambaut, A., and Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* 23, 7–9.

She, R. C., Hymas, W. C., Taggart, E. W., Petti, C. a, and Hillyard, D. R. (2010). Performance of enterovirus genotyping targeting the VP1 and VP2 regions on non-typeable isolates and patient specimens. *J Virol Methods* 165, 46–50.

Siafakas, N., Georgopoulou, A., Markoulatos, P., Spyrou, N., and Stanway, G. (2001). Molecular detection and identification of an enterovirus during an outbreak of aseptic meningitis. *J Clin Lab Anal* 15, 87–95.

Silva, P. A., Diedrich, S., de Paula Cardoso, D. D. D., and Schreier, E. (2008). Identification of enterovirus serotypes by pyrosequencing using multiple sequencing primers. *J Virol Methods* 148, 260–4.

Simmonds, P. (2006). Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. *J Virol* 80, 11124–40.

Simmonds, P. (2012). SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 5, 50.

Simmonds, P., McIntyre, C., Savolainen-Kopra, C., Tapparel, C., Mackay, I. M., and Hovi, T. (2010). Proposals for the classification of human rhinovirus species C into genotypically assigned types. *J Gen Virol* 91, 2409–19.

Simmonds, P., and Midgley, S. (2005). Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol* 79, 15467–76.

Simmonds, P., and Welch, J. (2006). Frequency and dynamics of recombination within different species of human enteroviruses. *J Virol* 80, 483–93.

Skern, T., and Sommergruber, W. (1985). Human rhinovirus 2: complete nucleotide sequence and proteolytic processing signals in the capsid protein region. *Nucleic Acids Res* 13, 2111–2126.

Skern, T., Torgersen, H., Auer, H., Kuechler, E., and Blaas, D. (1991). Human rhinovirus mutants resistant to low pH. *Virology* 183, 757–63.

Slater, A., Shann, F., and Pearson, G. (2003). PIM2: a revised version of the Paediatric Index of Mortality. *Intensive care medicine* 29, 278–85.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* 34, 126–9.

Smith, T. J., Kremer, M. J., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M. G., McKinlay, M. A., Diana, G. D., and Otto, M. J. (1986). The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science* 233, 1286–93.

Smura, T., Blomqvist, S., Paananen, A., Vuorinen, T., Sobotová, Z., Bubovica, V., Ivanova, O., Hovi, T., and Roivainen, M. (2007a). Enterovirus surveillance reveals proposed new serotypes and provides new insight into enterovirus 5'-untranslated region evolution. *J Gen Virol* 88, 2520–6.

Smura, T. P., Junttila, N., Blomqvist, S., Norder, H., Kaijalainen, S., Paananen, A., Magnius, L. O., Hovi, T., and Roivainen, M. (2007b). Enterovirus 94, a proposed new serotype in human enterovirus species D. *J Gen Virol* 88, 849–58.

Smuts, H. E., Workman, L. J., and Zar, H. J. (2011). Human rhinovirus infection in young African children with acute wheezing. *BMC Infect Dis* 11, 65.

Smyth, A., Smyth, R. L., Tong, C. Y., Hart, C. a, and Heaf, D. P. (1995). Effect of respiratory virus infections including rhinovirus on clinical status in cystic fibrosis. *Arch Dis Child* 73, 117–20.

Sommergruber, W., Casari, G., Fessl, F., Seipelt, J., and Skern, T. (1994). The 2A proteinase of human rhinovirus is a zinc containing enzyme. *Virology* 204, 815–8.

Spector, D. H. (1974). Requirement of 3'-Terminal Poly(adenylic Acid) for the Infectivity of Poliovirus RNA. *Proc Natl Acad Sci USA* 71, 2983–2987.

Stanway, G., Kalkkinen, N., Roivainen, M., Ghazi, F., Khan, M., Smyth, M., Meurman, O., and Hyypiä, T. (1994). Molecular and biological characteristics of echovirus 22, a representative of a new picornavirus group. *J Virol* 68, 8232–8.

Steinhauer, D. A., Domingo, E., and Holland, J. J. (1992). Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 122, 281–8.

Strimmer, K. (1994). Quartet Puzzling: A Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. *Mol Biol Evol* 13, 964–969.

Strimmer, K., and von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci USA* 94, 6815–9.

Suomalainen, M., and Greber, U. F. (2013). Uncoating of non-enveloped viruses. *Curr Opin Virol* 3, 27–33.

Swofford, D. L., Waddell, P. J., Huelsenbeck, J. P., Foster, P. G., Lewis, P. O., and Rogers, J. S. (2001). Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic biology* 50, 525–39.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24, 1596–9.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 101, 11030–5.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731–9.

Tapparel, C., Cordey, S., Van Belle, S., Turin, L., Lee, W.-M., Regamey, N., Meylan, P., Mühlemann, K., Gobbini, F., and Kaiser, L. (2009a). New molecular detection tools adapted to emerging rhinoviruses and enteroviruses. *J Clin Micro* 47, 1742–9.

Tapparel, C., Junier, T., Gerlach, D., Cordey, S., Van Belle, S., Perrin, L., Zdobnov, E. M., and Kaiser, L. (2007). New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features. *BMC Genomics* 8, 224.

Tapparel, C., Junier, T., Gerlach, D., Van-Belle, S., Turin, L., Cordey, S., Mühlemann, K., Regamey, N., Aubert, J.-D., Soccal, P. M., et al. (2009b). New respiratory enterovirus and recombinant rhinoviruses among circulating picornaviruses. *Emerg Infect Dis* 15, 719–26.

Tapparel, C., L'Huillier, A. G., Rougemont, A. L., Beghetti, M., Barazzone-Argiroffo, C., and Kaiser, L. (2009c). Pneumonia and pericarditis in a child with HRV-C infection: a case report. *J Clin Virol* 45, 157–60.

Taylor-Robinson, D., and Tyrrell, D. A. (1962). Serotypes of viruses (rhinoviruses) isolated from common colds. *Lancet* 1, 452–4.

Templeton, K. E. (2007). Why diagnose respiratory viral infection? *J Clin Virol* 40 Suppl 1, S2–4.

Templeton, K. E., Scheltinga, S. A., Beersma, M. F. C., Kroes, A. C. M., and Claas, E. C. J. (2004). Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza a and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. *J Clin Microbiol* 42, 1564–9.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673–80.

Toyoda, H., Nicklin, M. J., Murray, M. G., Anderson, C. W., Dunn, J. J., Studier, F. W., and Wimmer, E. (1986). A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell* 45, 761–70.

Tuthill, T. J., Groppelli, E., Hogle, J. M., and Rowlands, D. J. (2010). Picornaviruses. *Curr Top Microbiol Immunol* 343, 43–89.

Uncapher, C. R., DeWitt, C. M., and Colonno, R. J. (1991). The major and minor group receptor families contain all but one human rhinovirus serotype. *Virology* 180, 814–7.

Valcour, V., Haman, A., Cornes, S., Lawall, C., Parsa, A. T., Glaser, C., Yagi, S., Tihan, T., Bhatnagar, J., and Geschwind, M. (2008). A case of enteroviral meningoencephalitis presenting as rapidly progressive dementia. *Nat Clin Prac Neurol* 4, 399–403.

Venkataraman, S., Reddy, S. P., Loo, J., Idamakanti, N., Hallenbeck, P. L., and Reddy, V. S. (2008). Structure of Seneca Valley Virus-001: an oncolytic picornavirus representing a new genus. *Structure* 16, 1555–61.

Verdaguer, N., Blaas, D., and Fita, I. (2000). Structure of human rhinovirus serotype 2 (HRV2). *J Mol Biol* 300, 1179–94.

Victoria, J. G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S., and Delwart, E. (2009). Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol* 83, 4642–51.

Voss, T., Meyer, R., and Sommergruber, W. (1995). Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for the structural integrity. *Protein Sci* 4, 2526–31.

de Vries, M., Pyrc, K., Berkhout, R., Vermeulen-Oost, W., Dijkman, R., Jebbink, M. F., Bruisten, S., Berkhout, B., and van der Hoek, L. (2008). Human parechovirus type 1, 3, 4, 5, and 6 detection in picornavirus cultures. *J Clin Micro* 46, 759–62.

Véras, N. M. C., Gray, R. R., Brígido, L. F. D. M., Rodrigues, R., and Salemi, M. (2011). High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. *J Gen Virol* 92, 1698–709.

Wallis, C., and Melnick, J. L. (1967). Virus aggregation as the cause of the non-neutralizable persistent fraction. *J Virol* 1, 478–88.

Welch, J. B., McGowan, K., Searle, B., Gillon, J., Jarvis, L. M., and Simmonds, P. (2001). Detection of enterovirus viraemia in blood donors. *Vox sanguinis* 80, 211–5.

Whitton, J., Cornell, C., and Feuer, R. (2005). Host and virus determinants of picornavirus pathogenesis and tropism. *Nat Rev Micro* 3, 765–776.

Wiertsema, S. P., Chidlow, G. R., Kirkham, L. S., Corscadden, K. J., Mowe, E. N., Vijayasekaran, S., Coates, H. L., Harnett, G. B., and Richmond, P. C. (2011). High Detection Rates of Nucleic Acids of a Wide Range of Respiratory Viruses in the Nasopharynx and the Middle Ear of Children With a History of Recurrent Acute Otitis Media. *J Med Virol* 2017, 2008–2017.

Willcocks, M. M., Locker, N., Gomwalk, Z., Royall, E., Bakhshesh, M., Belsham, G. J., Idamakanti, N., Burroughs, K. D., Reddy, P. S., Hallenbeck, P. L., et al. (2011). Structural features of the Seneca Valley virus internal ribosome entry site (IRES) element: a picornavirus with a pestivirus-like IRES. *J Virol* 85, 4452–61.

Wisdom, A., Kutkowska, A. E., McWilliam Leitch, E. C., Gaunt, E., Templeton, K., Harvala, H., and Simmonds, P. (2009a). Genetics, recombination and clinical features of human rhinovirus species C (HRV-C) infections; interactions of HRV-C with other respiratory viruses. *PLoS One* 4, e8518.

Wisdom, A., McWilliam Leitch, E. C., Gaunt, E., Harvala, H., and Simmonds, P. (2009b). Screening respiratory samples for detection of human rhinoviruses (HRVs) and enteroviruses: comprehensive VP4-VP2 typing reveals high incidence and genetic diversity of HRV species C. *J Clin Micro* 47, 3958–67.

Witwer, C., Rauscher, S., Hofacker, I. L., and Stadler, P. F. (2001). Conserved RNA secondary structures in Picornaviridae genomes. *Nucleic Acids Res* 29, 5079–89.

Worobey, M., and Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J Gen Virol*, 2535–2543.

Xatzipsalti, M., Kyrana, S., Tsolia, M., Psarras, S., Bossios, A., Laza-Stanca, V., Johnston, S. L., and Papadopoulos, N. G. (2005). Rhinovirus viremia in children with respiratory infections. *Am J Respir Crit Care Med* 172, 1037–40.

Xiang, Z., Gonzalez, R., Wang, Z., Xiao, Y., Chen, L., Li, T., Vernet, G., Paranhos-Baccalà, G., Jin, Q., and Wang, J. (2010a). Human rhinoviruses in Chinese adults with acute respiratory tract infection. *J Infect* 61, 289–98.

Xiang, Z., Gonzalez, R., Xie, Z., Xiao, Y., Liu, J., Chen, L., Liu, C., Zhang, J., Ren, L., Vernet, G., et al. (2010b). Human rhinovirus C infections mirror those of human rhinovirus A in children with community-acquired pneumonia. *J Clin Virol* 49, 94–9.

Yozwiak, N. L., Skewes-Cox, P., Gordon, A., Saborio, S., Kuan, G., Balmaseda, A., Ganem, D., Harris, E., and DeRisi, J. L. (2010). Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. *J Virol* 84, 9047–58.

Zhang, Y., Zhu, S., Yan, D., Liu, G., Bai, R., Wang, D., Chen, L., Zhu, H., An, H., Kew, O., et al. (2010). Natural type 3/type 2 intertypic vaccine-related poliovirus recombinants with the first crossover sites within the VP1 capsid coding region. *PloS one* 5, e15300.

JOURNAL OF VIROLOGY, Oct. 2010, p. 10297–10310 Vol. 84, No. 19
0022-538X/10/$12.00 doi:10.1128/JVI.00962-10

Analysis of Genetic Diversity and Sites of Recombination in Human Rhinovirus Species C[▽]

Chloe L. McIntyre,[1] E. Carol McWilliam Leitch,[1] Carita Savolainen-Kopra,[2] Tapani Hovi,[2] and Peter Simmonds[1]*

Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, United Kingdom,[1] and National Institute for Health and Welfare, Mannerheimintie 166, FIN-00300 Helsinki, Finland[2]

Received 3 May 2010/Accepted 16 July 2010

Human rhinoviruses (HRVs) are a highly prevalent and diverse group of respiratory viruses. Although HRV-A and HRV-B are traditionally detected by virus isolation, a series of unculturable HRV variants have recently been described and assigned as a new species (HRV-C) within the picornavirus *Enterovirus* genus. To investigate their genetic diversity and occurrence of recombination, we have performed comprehensive phylogenetic analysis of sequences from the 5′ untranslated region (5′ UTR), VP4/VP2, VP1, and 3Dpol regions amplified from 89 HRV-C-positive respiratory samples and available published sequences. Branching orders of VP4/VP2, VP1, and 3Dpol trees were identical, consistent with the absence of intraspecies recombination in the coding regions. However, numerous tree topology changes were apparent in the 5′ UTR, where >60% of analyzed HRV-C variants showed recombination with species A sequences. Two recombination hot spots in stem-loop 5 and the polypyrimidine tract in the 5′ UTR were mapped using the program GroupingScan. Available HRV-C sequences showed evidence for additional interspecies recombination with HRV-A in the 2A gene, with breakpoints mapping precisely to the boundaries of the C-terminal domain of the encoded proteinase. Pairwise distances between HRV-C variants in VP1 and VP4/VP2 regions fell into two separate distributions, resembling inter- and intraserotype distances of species A and B. These observations suggest that, without serological cross-neutralization data, HRV-C genetic groups may be equivalently classified into types using divergence thresholds derived from distance distributions. The extensive sequence data from multiple genome regions of HRV-C and analyses of recombination in the current study will assist future formulation of consensus criteria for HRV-C type assignment and identification.

A series of studies published in 2006 to 2007 described a novel group of human rhinoviruses (HRVs), referred to as A2 or group C (HRV-C), genetically distinct from existing rhinoviruses (3, 21, 23, 25, 27, 33, 42). These and subsequent investigations have shown the proposed new species HRV-C to be remarkably prevalent, widely distributed geographically, and frequently associated with severe respiratory disease, both following primary infections, particularly in young children (24, 25, 29, 32, 59), and as an exacerbating factor in asthma and other chronic obstructive airway diseases (20, 25, 29, 36). On the other hand, HRV-C has also been abundant in specimens collected from both children and adults with milder cases of respiratory disease (47, 48).

The HRV-C species is a member of the *Enterovirus* genus within the family *Picornaviridae* (53) and shares many features of its genome organization and structure with other picornaviruses and, more specifically, with other members of the *Enterovirus* genus. This includes an approximately 7,100-base genome containing a single reading frame; this encodes four capsid proteins and a series of functionally conserved nonstructural proteins involved in virus replication. HRV-C, in common with other members of the *Enterovirus* genus, possesses a type I internal ribosomal entry site (IRES) that enables initiation of translation from an internal methionine codon several hundreds of bases from the 5′ end of the genome.

Although no more closely related to each other than they are to other *Enterovirus* species phylogenetically, the three species of human rhinoviruses (A, B, and C) share a number of biological and genetic attributes. Most prominent is their transmission route and primary tropism for the respiratory tract, associated with the known acid lability of species A and B that is traditionally believed to prevent their colonization of the gastrointestinal tract. The HRV genome possesses a lower G+C content than genomes of other enteroviruses, and it is often speculated that this represents an adaptation for replication at the lower temperatures found in upper airways. HRV-C differs from other species, however, in other aspects, including the siting of the *cis*-replicating element in VP2 (11) instead of in 2A (species A) (13) or VP1 (species B) (34). Species C has so far proven refractory to all attempts at *in vitro* culture, despite the use of a wide variety of cell lines and primary cell cultures (25, 33). This restriction has hindered investigations of its replication, receptor use, and antigenic diversity.

Like most picornavirus groups, human rhinoviruses show substantial genetic heterogeneity that underlies the existence of a large number of antigenically distinct variants. In the case of species A and B, a total of 74 and 25 different serotypes, respectively, have been defined using cross-neutralization assays in cell culture (14, 18, 19). Available nucleotide sequences from the capsid region of HRV-C reveal similar or even greater diversity than is found between species A and B sero-

* Corresponding author. Mailing address: Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, United Kingdom. Phone: 44 131 650 7927. Fax: 44 131 650 6511. E-mail: Peter.Simmonds@ed.ac.uk.

▽ Published ahead of print on 28 July 2010.

types, but its lack of culturability currently precludes an equivalent classification of HRV-C into serotypes.

Although coding regions of HRV-C are genetically distinct from those of species A and B rhinoviruses and other enteroviruses, both the initial detection and subsequent genetic analyses of HRV-C have been complicated by the similarity of most (>60%) sequences of the 5′ untranslated region (UTR) to those of species A (49), with the remainder being phylogenetically distinct. It has been proposed that one or more interspecies recombination events between species C and species A rhinoviruses have occurred to generate these chimeric sequences (15, 17, 49, 59). Despite these differences in genetic composition, no differences in clinical presentations or epidemiology have been detected between HRV-C variants with species A-like and species C-like 5′ UTR (HRV-Ca and HRV-Cc, respectively) sequences (17, 59).

In the current study we have generated comparative sequence data from several genome regions (5′ UTR, VP4/VP2, VP1, and 3Dpol) from a large number of HRV-C-positive clinical specimens. Combined with published sequence information from these regions, we have investigated the occurrence, frequency, and location of recombination events within this rhinovirus species. The information gained on comparative phylogenies and sequence divergence in different regions provides data that will assist in the eventual classification of the current plethora of HRV-C sequences into a number of genetically determined types.

MATERIALS AND METHODS

Genetic characterization of HRV-C variants. A total of 89 HRV-C-positive samples from the Edinburgh Specialist Virology Centre (SVC) respiratory sample archive (60) were selected for genetic characterization; these represent a subset of the 144 samples previously characterized in the VP4/partial VP2 region (59). RNA was extracted from clinical samples as previously described (60). Sequences were amplified from four different genomic regions using the methods described below.

(i) VP4/VP2. The VP4 and 5′ end of VP2 were amplified using nested primers (see Table S1 posted at http://www.virus-evolution.org/Downloads/JVI00962-10/) as previously described (59).

(ii) VP1. Nested reverse transcription-PCRs (RT-PCRs) were performed as previously described (35) but using primers shown in Table S1 (posted at the URL mentioned above). A total of 72 from the 89 samples amplifiable in other genome regions were successfully amplified in two VP1 regions. The two sequence fragments overlapped by 105 bases, allowing a composite 777-base sequence to be generated for phylogenetic analysis.

(iii) 3Dpol and 5′ UTR. RNA extracted from respiratory samples was reverse transcribed into cDNA using the Promega reverse transcription system (Promega, United Kingdom) as per the manufacturer's instructions except that 5 μl of extracted RNA template was used. Amplification of the 3Dpol region and the 5′ UTR region from cDNA used nested primers (see Table S1 posted at the URL mentioned above) and previously described PCR conditions (59) but with an annealing temperature of 48°C in the second round. Samples negative on initial amplification were amplified using the nested PCR strategy described for VP1 above. A 3Dpol amplicon of 470 nucleotides in length was produced. For the 5′ UTR, two sequence fragments with a 61-base overlap allowed a 680-base composite 5′ UTR sequence overlapping the VP4/VP2 amplicon to be generated. A subset of HRV-C variants ($n = 12$) was further amplified using heminested, combined species A and C primers (5′ UTR set 1 in Table S1, posted at http://www.virus-evolution.org/Downloads/JVI00962-10/) to generate an amplicon from nucleotide position 1 to 355 (27 to 335, excluding primer sequences).

All amplicons were sequenced using a BigDye Terminator kit (Applied Biosystems, Warrington, United Kingdom). Prior to sequencing reactions, all samples underwent a PCR product cleanup procedure using EXOSAP-IT (GE Healthcare, United Kingdom).

Sequence analysis. All available VP4/VP2 region sequences were downloaded from GenBank on 26 April 2010. Sequences that were <90% complete across the region from position 615 to 1043 were discarded (sequence positions are numbered throughout the manuscript using the sequence 024 [accession number EF582385]; the current GenBank entry for the QPM prototype sequence [EF186077] is incomplete at the 5′ end). These comprised 538 sequences along with a further 16 previously unpublished sequences from the SVC archive. The sequence data set was supplemented with the 10 complete genome sequences of HRV-C currently available on public databases as well as all available complete genome sequences of species A ($n = 82$) and B ($n = 25$) of defined serotypes.

Phylogenetic trees were constructed using the MEGA package, version 4.0 (54), by the neighbor-joining method (44) from 100 bootstrap resampled sequence alignments of maximum composite likelihood distances (MCL) (55) with pairwise deletion for missing data. Distributions of pairwise sequence distances for identification of type thresholds were calculated using the program SequenceDist in the Simmonics package.

Recombination breakpoints in HRV-Ca 5′ UTR sequences were determined using the program GroupingScan (51), using complete genome sequences of species A ($n = 82$), B ($n = 25$), and nonrecombinant species C (Cc) sequences ($n = 29$) as control groups. Trees were constructed using a fragment size of 200 with incremental steps of 10 bases, a bootstrap value of 70% to define phylogenetically supported groups, and 10 sequence relabelings (default) to calculate control values. A second analysis of sequences extending to the extreme 5′ end of the genome was performed to investigate further recombination events around the replication structures (eight HRV-Cc control sequences). Recombination breakpoints were interpolated as the nucleotide positions where the grouping values switched from the A group to the C group.

5′ UTR RNA structure prediction. Structure prediction for HRV-C was based on the established model for HRV-A (6) and previous comparative analyses of HRV-A and -B sequences (43, 56). Mfold analysis of a selection of HRV-Ca and HRV-Cc sequences produced minimum-energy secondary structures for HRV-C variants that closely matched the HRV-A2 structure (data not shown).

Nucleotide sequence accession numbers. All newly generated sequences were submitted to GenBank and were assigned accession numbers HM236897 to HM236968 (VP1), HM352737 to HM352752 (VP4), HM485468 to HM485556 (3Dpol), and HM581802 to HM581888 (5′ UTR).

RESULTS

Phylogenies of HRV-C 5′ UTR, VP4/VP2, VP1, and 3Dpol regions. To investigate sequence relationships of HRV-C in different regions of the genome, 89 study samples were sequenced in the 5′ UTR (nucleotide position 185 to 614; $n = 87$), VP4 and partial VP2 (615 to 1043), whole VP1 (2348 to 3125; $n = 72$), and partial 3Dpol regions (6384 to 6854) and aligned with the 10 available complete genome sequences of HRV-C (Fig. 1). On phylogenetic analysis, sequences in each region fell into a number of clearly defined, bootstrap-supported phylogenetic groups, with the majority containing several variants and others represented by single sequences. The small subset of variants for which we were unable to obtain VP1 sequences were scattered throughout the 5′ UTR, VP4/VP2, and 3Dpol trees. They therefore did not constitute an identifiable genetic subset of sequences whose omission would have biased the analysis of VP1 sequence diversity.

The phylogenies of the three coding regions (VP4/VP2, VP1, and 3Dpol) were remarkably congruent, with identical grouping patterns for all but one of the sequences analyzed. Using a bootstrap value of 70% or greater to define clades, there was only one phylogeny violation between trees constructed from coding region sequences (Fig. 1B, C, and D). This involved the previously characterized complete genome sequence, N4 (GQ223227) (Fig. 1, shaded box), which changed tree position between VP1 and 3Dpol regions. Our sample collection contained a single variant (Resp_10300/08) that grouped with N4 in the 5′ UTR, VP4/VP2, and VP1 regions. In contrast to N4, its 3Dpol sequence maintained a similar phylogenetic position relative to other sequences, as was observed

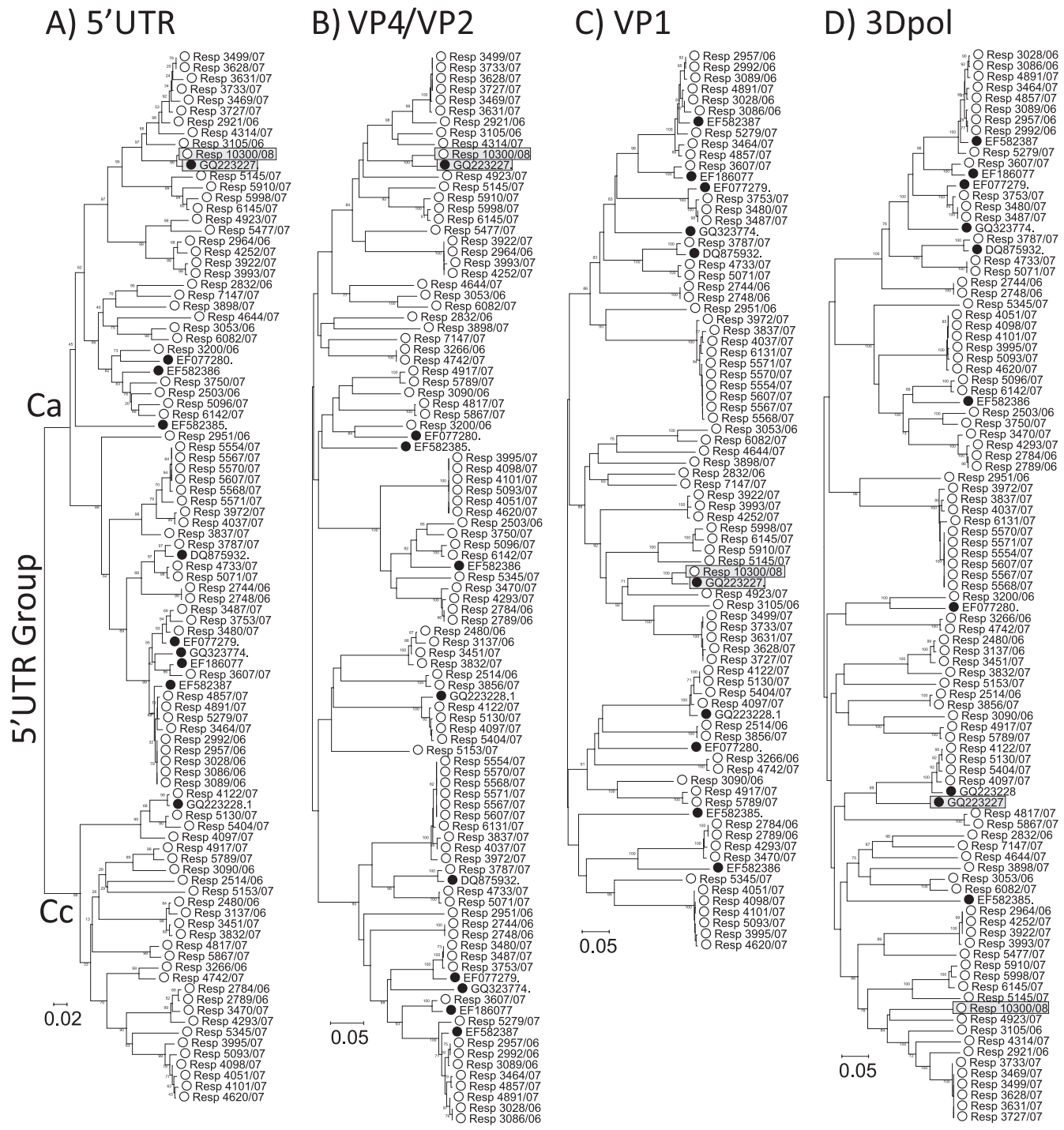


FIG. 1. Phylogenetic analysis of study sequences and corresponding regions from complete genome sequences in the 5′ UTR (position 185 to 614, numbered according to the reference sequence EF582385) (A), VP4/VP2 (616 to 1043) (B), VP1 (2348 to 3125) (C), and 3Dpol (6384 to 6854) (D). Trees were constructed by neighbor joining of pairwise maximum composite likelihood distance implemented in the program MEGA (54); branches showing at least 70% bootstrap support are indicated. Complete genome sequences were labeled using filled symbols. The putative recombinant sequence N4 (GQ223227) and the other group 7 sequence Resp_10300/08 are indicated by shaded boxes.

in the 5′ UTR, VP4/VP2, and VP1 regions (Fig. 1). In investigating this anomaly, we found on closer inspection of the N4 sequence in the 3Dpol region that its 5′ end (position 6384 to 6645) was similar to that of Resp_10300/08 (pairwise distance, 0.08), while its 3′ end (position 6646 to 6854) was nearly identical to another sequence, N10 (GQ223228), originating from the same laboratory (sequence divergence, 0.01, compared to 0.35 divergence from Resp_10300/08). The chimeric nature of this sequence is evident from the change in tree position of 5′- and 3′-end fragments of the 3Dpol region (see Fig. S1 and

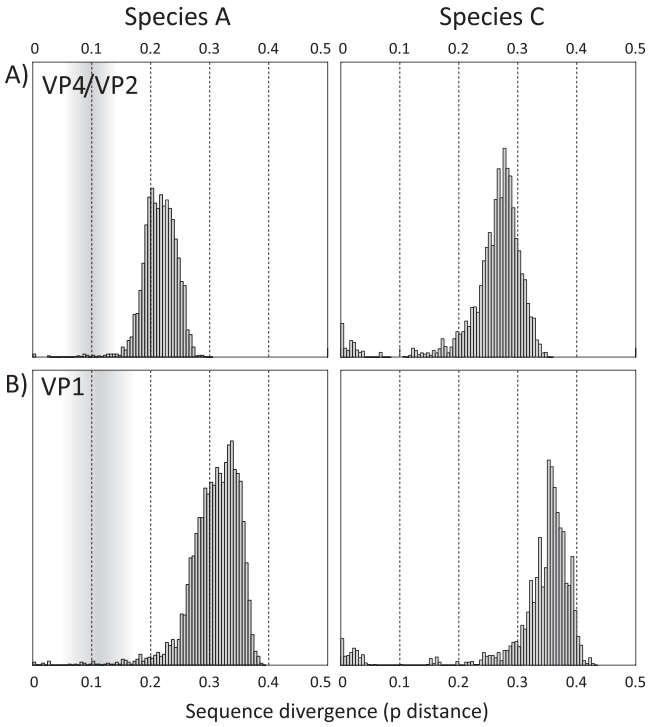


FIG. 2. Distribution of pairwise *p* distances between HRV-A sequences (A) and available sequences of study samples and complete genome sequences of HRV-C (B) in the VP4 and VP1 regions. The shaded area represents the zone between inter- and intraserotype divergence values in species A.

supplementary data posted at http://www.virus-evolution.org/Downloads/JVI00962-10/); N4 grouped with Resp_10300/08 at the 5′ half of the fragment (comparable to other genome regions) (Fig. 1B and C) but clustered very closely with N10 (GQ223228) at the 3′ end. Possible underlying explanations for the anomalous phylogeny of the N4 sequence in the 3Dpol region are discussed below.

In contrast to the identity of coding region phylogenies, a minimum of 12 phylogeny violations were observed in the species C sequence data set on comparison of phylogenetic trees from the 5′ UTR and VP4/VP2. Members of the previously described HRV-Cc 5′ UTR group, distinct from the species A-like 5′ UTR sequences of the majority of HRV-C sequences (labeled in Fig. 1), were drawn from several coding region lineages. These observations are consistent with the occurrence of multiple recombination events between the 5′ UTR and the rest of the genome.

Pairwise distance distributions. Distributions of pairwise distances between HRV-C sequences in VP4/VP2 and VP1 regions were computed to investigate whether they contained clear sequence divergence thresholds that might be used to define intertype and intratype divergence values (Fig. 2B). These distributions were compared with those of species A in the same genome regions (Fig. 2A). They were also similar to those of species B, but in the latter case there were insufficient numbers of classified sequences in either genome region for a conclusive analysis.

The set of HRV-C VP4/VP2 and VP1 pairwise distance values (Fig. 2B) contained a first distribution (maximum values of 6.9% and 8%, respectively) that corresponded to pairwise distances between variants within the same phylogenetic clades (Fig. 1B and C). These were comparable to the majority of (intraserotype) distances between the small number of available HRV-A sequences of the same serotype (Fig. 2A). A second, nonoverlapping distribution, with minimum values of 10.9% in VP4 and 14.7% in VP1, corresponded to the set of pairwise distances between HRV-C clades. There was a comparable division of pairwise distances between VP4/VP2 and VP1 sequences of species A (Fig. 2A) and species B (data not shown) into two largely separate distributions. The previously proposed VP1 divergence thresholds of 12% and 13% for serotype assignment (22) correspond closely to the low point in the distribution of HRV-A and HRV-B VP1 pairwise distances. With the currently available sequence data, a threshold around this divergence value in VP1 and 10% in VP4/VP2 would appear appropriate for HRV-C. Observations of the separate distributions of pairwise distances in VP1 and phylogenetic evidence for marked phylogenetic clustering of HRV-C in all three coding regions analyzed support the idea that HRV-C variants might be usefully classified into a number of (genetically determined) groups that correspond to serotypes of other rhinoviruses. Such assignments might be made even though we lack serological evidence for antigenic distinctiveness of HRV-C variants.

As a preparatory step toward a future genotypic classification of HRV-C, we have tabulated all available HRV-C variants into groups containing members that show >0.13 sequence divergence from each other in VP1 or, if not available, >0.10 in VP4/VP2 (Table 1). Table 1 has been split into two sections, one for which complete genome sequences are available (genetic groups numbered 1 to 11) and a second containing genetic groups for which VP4/VP2 sequences are available, with or without VP1, 3Dpol, and 5′ UTR sequences. Combining the assignments, if sequences grouped in this way are considered the equivalent of (sero)types in other rhinoviruses, then the current data set of sequences can be grouped into a total of 60 genetically determined types.

Sequence divergence of HRV-C in different genome regions. Using the genetic groups in the previous section, we compared the degree of sequence divergence in the analyzed genome regions with divergence in regions of different serotypes of HRV-A and HRV-B species (Fig. 3). Both at the nucleotide and amino acid level, VP1 was the most variable region between types and serotypes, followed by 3Dpol and VP4/VP2. However, HRV-C genetic groups were consistently more diverse than serotypes of other rhinovirus species, showing, for example, 35% (nucleotide) and 33% (amino acid) divergence compared to 25% and 21 to 26% in HRV-A and -B serotypes. Even more marked differences in 3Dpol and VP4/VP2 regions were observed, with amino acid divergence between species C groups almost twice that of species A and B.

Mapping positions of recombination between the 5′ UTR and VP4/VP2. Combining sequences from overlapping 5′ UTR and VP4/VP2 amplicons provided a continuous sequence from position 185 to 1043 that enabled the recombination breakpoint(s) to be determined in each of the Ca recombinant variants identified to date. The combined set of 10 available complete genomes, 87 samples from the current study, and 34 5′

TABLE 1. Division of HRV-C into genetic groups based on divergence in VP1 or VP4/VP2

HRV-C genetic group[a]	Accession no.[b]	Identifier	5′ UTR group[c]	Submission date (mo/day/yr)[d]	Reference[e]
Groups containing complete genomes					
1	EF077279	NAT001	Ca	10/20/06	21
2	EF077280	NAT045	Ca	10/20/06	21
3	EF186077	QPM	Ca	12/14/06	33
4	EF582385	024	Ca	4/27/07	25
5	EF582386	025	Ca	4/27/07	25
6	EF582387	026	Ca	4/27/07	25
7	DQ875932	NY-074	Ca	7/14/08	23
8	GQ223227	N4	Ca	5/29/09	17
9	GQ223228	N10	Cc	5/29/09	17
10	GQ323774	QCE	Ca	6/29/09	2
11	FJ392317[f]	CL170085	Ca	10/17/08	58
Groups containing partial genomes					
12	EF077264	NAT083	Ca	10/20/06	21
13	EU081795	tu403	Ca	8/03/07	42
14	EU081796	06-445	Cc	8/03/07	42
15	EU081800	06-20	Ca	8/03/07	42
16	EU081808	g2-4	Ca	8/03/07	42
17	EU081809	06-582	Ca	8/03/07	42
18	EU590074	PNC41788	Ca	3/25/08	48
19	EU697850	7316563	Cc	5/05/08	7
20	EU697851	DK-1	Cc	5/05/08	7
21	EU752377	RV471	Ca	5/26/08	36
22	EU752381	RV541	Ca	5/26/08	36
23	EU752424	RV64	Ca	5/26/08	36
24	EU752426	RV408	Ca	5/26/08	36
25	EU752427	RV1123	Ca	5/26/08	36
26	EU752441	RV177	Ca/Cc[g]	5/26/08	36
27	GQ223122	N22	Ca/Cc	1/09/09	17
28	GQ223134	N46	Ca	1/09/09	17
29	FJ615699	201882	Cc	1/09/09	37
30	GQ476669	Resp_3898	Ca	8/13/09	60
31	GU294380	Resp_3776	Ca	12/04/09	59
32	GU294466	Resp_5613	Ca	12/04/09	59
33	GU294480	Resp_6157	Cc	12/04/09	59
34	EF077256	NAT069	Cc	10/20/06	21
35	EF077260	NAT069	Cc	10/20/06	21
36	EU081790	06-739	Cc	8/03/07	42
37	EU081791	tu304	Ca	8/03/07	42
38	EU081799	g2-11		8/03/07	42
39	EU081802	g2-25	Ca	8/03/07	42
40	EU081803	g2-23		8/03/07	42
41	EU081805	g2-28	Ca	8/03/07	42
42	EU081807	06-230		8/03/07	42
43	EU590054	PNC86718	Cc	3/25/08	48
44	EU590061	PNC40168		3/25/08	48
45	EU590064	PNC40449		3/25/08	48
46	EU697839	IN-36		5/05/08	7
47	EU697852	SA365412		5/05/08	7
48	EU743925	CO-1368	Cc	5/22/08	11a
49	EU752358	RV1250		5/26/08	36
50	EU752398	RV1039		5/26/08	36
51	EU752412	RV546	Cc	5/26/08	36
52	FJ598096	GDFY100	Cc	12/29/08	Unpublished
53	FJ615722	202511		1/09/09	37
54	FJ615737	202092		1/09/09	37
55	FJ615745	202642		1/09/09	37
56	FJ841957	S05986		3/17/09	9
57	FJ869923	KR1868		3/27/09	15
58	FJ869950	KR2315	Cc	3/27/09	15
59	GQ466482	K1091 301104		8/07/09	47
60	GU214340	PV68	Ca	11/18/09	41

[a] Genetic groups showing >0.10 sequence divergence in VP4/VP2. Groups for which VP4/VP2 sequences are available, with or without VP1, 3DPol, and 5′ UTR sequences, were considered partial.

[b] Accession number of the first member of the assigned genetic group.

[c] Identity of the 5′ UTR group (Ca or Cc) associated with viruses of this group (59).

[d] Submission date of the first sequence of the assigned genetic group in GenBank/EMBL/DDBJ.

[e] Citation for the first submitted member of the genetic group.

[f] Accession number FJ392317 refers to a previously published VP1 sequence (58). The full genome of this variant has now been completed (C. Tapparel, personal communication). In order to preserve the numbering of genetic groups, sequences grouping with this complete genome in VP4/VP2 have been assigned as genetic group 11.

[g] Variants with Ca and Cc 5′ UTR sequences are found in these genetic groups (59).

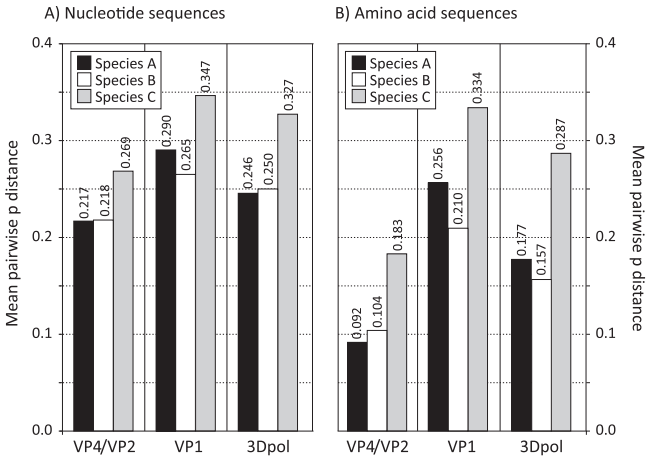


FIG. 3. Mean pairwise (uncorrected) *p* distances of nucleotide (A) and amino acid (B) sequences in the VP4/VP2, VP1, and 3Dpol regions between serotypes of HRV-A and HRV-B and between identified genetic groups of HRV-C (values shown above each bar).

UTR sequences described by Huang et al. (17) represented 45 of the 60 sequence groups listed in Table 1. Twenty-nine HRV-C sequence groups had 5′ UTR sequences that grouped within the species A 5′ UTR clade (labeled Ca in Fig. 1), two contained sequences from both the Ca and Cc groups, and the remainder were phylogenetically distinct, falling in the Cc clade that contained the N10 complete genome sequence (GQ223228).

The program Grouping Scan (51) was used to identify recombination breakpoints in the Ca 5′ UTR sequences. This program was chosen in preference to bootscanning methods because it scores the extent of grouping within predefined control groups rather than simply the bootstrap support for the grouping of a query sequence with a group consensus sequence. As discussed previously (51), the latter method can lead to false assignment in cases where a query sequence is not closely affiliated to any of the control groups. Furthermore, by simply condensing the often large amount of comparative sequence data within control groups to a single consensus sequence, bootscanning additionally discards informative data on sequence diversity within groups that are of value in assessing phylogeny relationships. The available set of HRV-A and HRV-B complete genome sequences were used as control groups while the HRV-Cc control group (nonrecombinant Cc sequences) was assembled from the N10 sequence and Cc variants identified in the current study ($n = 29$).

Each of the complete genome and study sequences from the Ca 5′ UTR clade was scanned between position 185 and 903 (three representative results depicting different breakpoints are shown in Fig. 4A). Superimposed is a diagrammatic summary of the RNA secondary structure elements comprising the replication structures (stem-loop 1 cloverleaf) and IRES (stem-loops 2 to 6, the latter containing the AUG start codon of the HRV-C open reading frame at position 615).

Recombination breakpoints identified in HRV-Ca variants were invariably positioned similarly among variants in the same genetic group (Table 2). Two recombination hot spots were identified, one occurring in the polypyrimidine tract

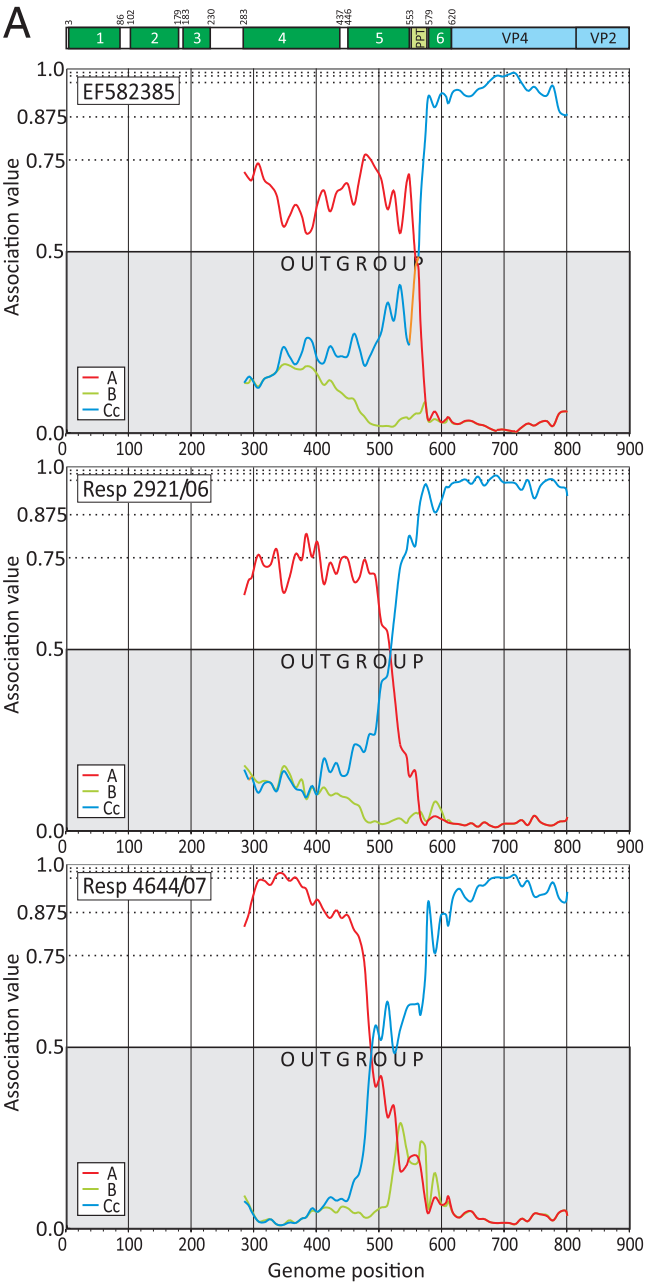


FIG. 4. (A) Mapping of recombination breakpoints of three representative published or study HRV-Ca sequences using the program Grouping Scan with HRV-A, -B, and HRV-Cc sequences as control groups, as indicated on the figure. Above the graphs is a diagrammatic summary of 5′ UTR RNA structure elements (stem-loops 1 to 6 [1] and the PPT). (B) Variability scan of the 5′ UTR, showing mean pairwise interserotype/genetic group distances of HRV-A, -B, and -Cc sequences averaged over a 30-base window and, separately, the interspecies divergence between HRV-A and HRV-Cc (A-Cc). The position of 5′ UTR structures and range of identified recombination breakpoints (Table 2) are indicated. (C) Extension of Grouping Scan analysis to the 5′ end of the HRV genome using the subset of control sequences complete from position 27 onwards and an HRV-Ca query sequence.

(PPT) between stem-loops 5 and 6 (position 561 to 576; mean, 565) and a second around the terminal loop of stem-loop 5 (position 508 to 544; mean, 523), while two HRV-C genetic groups showed a more 5′ site (position 479 to 489; mean, 481).

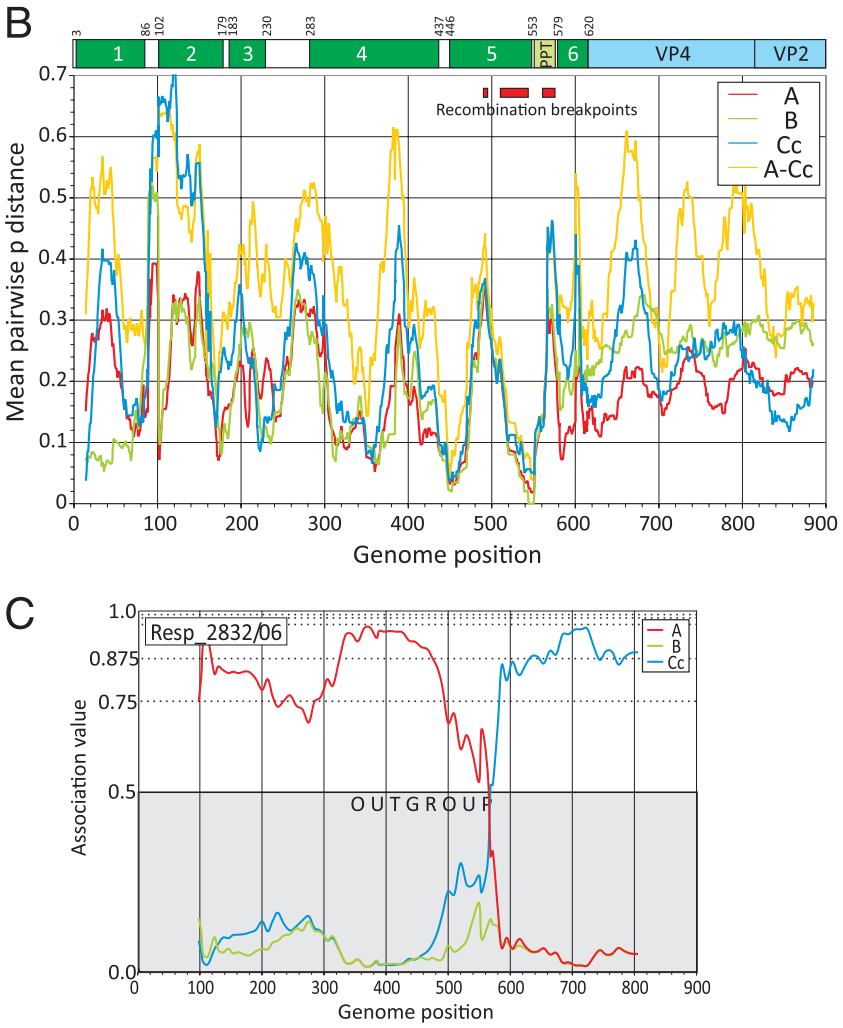


FIG. 4—*Continued.*

Sequences of the stem-loop stem regions either side of the recombination hot spot centering on position 523 were highly conserved (Fig. 4B). The HRV-Ca recombinants with a 5′ stem sequence derived from species A and a 3′ sequence from species C showed predicted pairings similar to those of HRV-A and nonrecombinant HRV-Cc sequences (data not shown) and would therefore accommodate a recombination event without destabilizing stem-loop 5.

Finally, the occurrence of further recombination events at the extreme 5′ end of the 5′ UTR was investigated using a subset of 11 HRV-Ca sequences with sequences complete from position 27. All query sequences grouped within the HRV-A clade, with association values substantially above the 0.5 outgroup position score (an example is shown in Fig. 4C). Although the resolution of this method is limited by the 200-base fragment size used for analysis, visual inspection of trees constructed from shorter sequence fragments (50 and 100 bases) at the extreme 5′ end of the genome provides no evidence for a change in the phylogenetic position of any Ca variants to a position outside the HRV-A clade (data not shown).

Interspecies recombination in the 2A region. To investigate whether other genome regions of HRV-C showed evidence for interspecies recombination, we performed a scan of sequence divergence between complete genome sequences and those of species A and B rhinoviruses (Fig. 5). Between-species divergence was substantially greater than the mean pairwise distance within species C sequences except for an approximately 260-base region in 2A. This region (position 3268 to 3525) showed a markedly different phylogeny from flanking regions of the same length (Fig. 6). More specifically, species C sequences grouped within the HRV-A clade in a manner similar to that observed in the 5′ UTR. As found in the latter region, single HRV-C sequences or groups of HRV-C sequences became scattered among the HRV-A sequences, implying the occurrence of separate recombination events. Remarkably, the bootstrap-supported grouping of five HRV-C sequences (accession numbers EF077279, DQ875932, EF582387, EF186077, and GQ323774) in 2A was similarly monophyletic in the 5′ UTR (Fig. 1A). The main difference from the 5′ UTR phylogeny was the position of N10 (GQ223228). This formed part of the Cc 5′ UTR clade (Fig. 1A) but contained a 2A sequence embedded within the HRV-A clade (Fig. 6). A change in the phylogenetic position of N4 (GQ223227) was additionally apparent in trees from the two regions.

TABLE 2. Position of recombination breakpoints in different HRV-Ca groups

Genetic group	Total no. of sequences analyzed	Breakpoint (mean nucleotide position [range])
1	5	520 (519–523)
2	1	564
3	2	526 (526–527)
4	1	561
5	3	563 (563–564)
6	10	524 (524–526)
7	2	574 (574)
8	4	565 (561–570)
10	1	525
11	2	554 (552–557)
12	4	508 (508)
13	1	523
15	3	483 (479–489)
16	3	520 (518–526)
17	1	567
18	8	528 (518–544)
21	6	572 (570–575)
22	2	576 (576)
23	3	562 (552–570)
24	1	560
25	1	558
27	3	535 (522–559)
28	2	543 (536–550)
30	1	561
31	1	528
32	9	524 (523–527)
37	2	555 (555)
39	1	560
41	1	479
60	1	516

DISCUSSION

Recombination in HRV-C. This study has investigated the occurrence and sites of recombination within the genomes of the recently described species C human rhinoviruses. The most striking finding was the almost identical phylogenetic trees of the three coding regions analyzed, including gene regions at the extreme 5′ and 3′ ends of the open reading frame (VP4/VP2 and 3Dpol). The absence of recombination implied by these observations is consistent with previous comparisons of VP4/VP2 and 3Dpol region phylogenies of HRV species A and B (46), where only minimal changes in branching order were observed among the full set of 101 classified serotypes. These findings together with the data obtained here for species C contrast dramatically with the rampant and ongoing recombination process between structural and nonstructural gene regions in other species within the *Enterovirus* genus, particularly in species A, B, and C human enteroviruses (10, 28, 30, 38, 52). Recombination is also extensively documented in aphthoviruses (16, 50), cardioviruses (5, 12), and parechoviruses (4, 8). For these, the concept of separate, modular evolution of structural and nonstructural regions of picornavirus genomes has been developed (28, 31, 50).

Potentially underlying this difference in recombination frequency in rhinoviruses is their different pattern of sequence divergence in structural and nonstructural regions from most other picornaviruses. Species C variants along with HRV-A and HRV-B show substantial sequence divergence throughout the coding region. For example, mean intergenetic group pairwise distances in the 3Dpol region of HRV-C (33% nucleotide; 29% amino acid) were similar to the divergence of VP1 (35% and 33%, respectively). This contrasts markedly with human enteroviruses and several other picornavirus groups where nonstructural gene regions are much less divergent between serotypes (<10% at the amino acid level). As previously proposed (50), this restricted variability increases the likelihood of recombinants with breakpoints in the nonstructural regions being viable biologically. In contrast, the highly divergent sequences in equivalent regions of HRV may be functionally incompatible with each other and effectively isolate each rhinovirus serotype into a separate evolutionary path.

One HRV-C sequence showed different phylogenetic relationships in different regions of the coding region (N4; GQ223227). Although we cannot exclude the possibility of a natural recombination site within the 3D region as an explanation for the hybrid nature of its 3Dpol sequence (see Fig. S1 and the supplementary data posted at http://www.virus-evolution.org/Downloads/JVI00962-10/), its close resemblance of the 3′ end to the N10 sequence originating from the same laboratory suggests the possibility of laboratory contamination during assembly of the complete genome sequence. Another group 7 variant (Resp_10300/08) showed consistent phylogenetic relationships to other sequences in all three coding regions analyzed, including 3Dpol (Fig. 1B, C, and D). It would be of value if the laboratory from which the N4 sequence originated (17) were able to perform additional amplification and sequencing to rule out contamination/assembly errors and correct the GenBank entry if necessary.

Evidence for a different, likely more evolutionarily ancient, pattern of recombination in HRV-A has been previously obtained, manifested by differences in sequence relatedness between sequences in different genome regions (39, 57). For example, HRV-53 was substantially more similar to HRV-46 in the nonstructural region than anticipated by their sequence relationship in the capsid-encoding region; conversely HRV-78 and HRV-12 were more divergent. In these and other specific examples, changes in the phylogeny relationship usually occurred at the P1/P2 boundary, implying a limited degree of compatibility between structural and nonstructural gene modules derived from different serotypes. As described above, this pattern of recombination is quite distinct from the multiple recombination sites within nonstructural gene regions of enteroviruses and other picornaviruses. No evidence for its occurrence was documented among the data set of HRV-C sequences assembled in the current study, where both small-scale (genetic group) and larger sequence groupings were preserved across the genome. Potentially, the much greater amino acid sequence divergence between HRV-C sequence groups (Fig. 3) increases the likelihood of biological incompatibility and restricts further the occurrence of recombination.

Despite the limited evidence for recombination in the coding regions of rhinoviruses, this study confirmed its frequent occurrence between VP4 (and the rest of the genome downstream) and the 5′ UTR (17, 49). The generation of a large number of 5′ UTR Ca and Cc sequences permitted a detailed investigation of the positions where recombination occurred. The distribution of recombination hot spots centered around position 565 (in the PPT), 523 (within stem-loop 5 of the

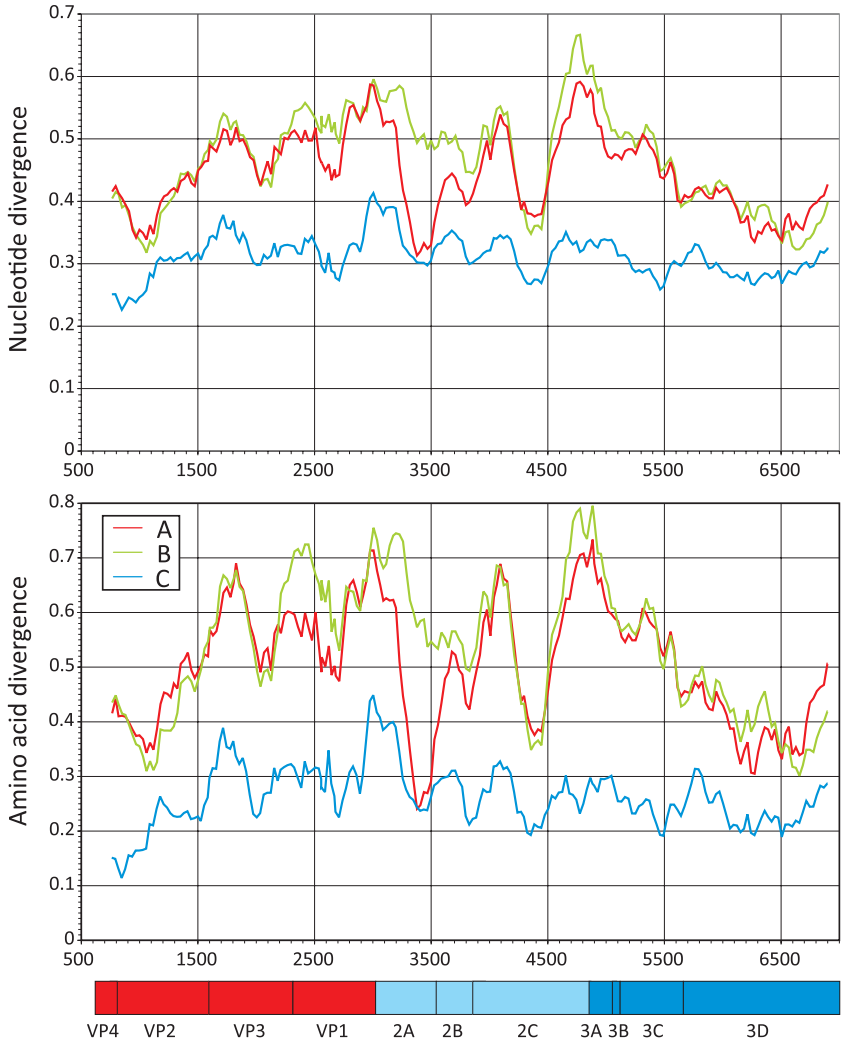


FIG. 5. Divergence scan, showing mean nucleotide and amino acid sequence divergence between complete genome sequences of HRV-C with those of HRV-A and -B and within-species diversity of HRV-C. The scan used fragment lengths of 300 nucleotides, incrementing by 30 bases across the HRV coding sequence. The genome diagram below the graph shows positions of different HRV-C genes, using annotation supplied with the sequence EF582385.

IRES), and two groups showing evidence of a breakpoint around position 481. This is a much more restricted distribution than previously described in an analysis using a smaller data set and a different scanning method (17). Nor could we confirm the second recombination event at the extreme 5′ end of the genome where Ca variants remained grouped with HRV-A sequences.

As suggested from the analysis of recombination in the coding regions, the occurrence and positions of recombination events in the 5′ UTR may be governed by biological compatibility restrictions. For example, the conservation of sequences within stem-loop 5 and potentially independent modular functions of stem-loops 5 and 6 may favor the creation of viable recombinants. Similarly, the high degree of sequence conservation of the 5′ end of the genome between rhinovirus species may facilitate the interaction between the species C replication complex and the HRV-A-derived cloverleaf replication structure (stem-loop 1) (56). In the future, *in vitro* insertion of the HRV-C-derived IRES and adjacent coding sequences into species A replicons or infectious clones will be of value in functionally mapping these compatibility restrictions.

A new and unexpected finding in the current study was the second site of HRV-A/HRV-C interspecies recombination in the 2A region (Fig. 6 and 7). Although a previous bootscanning analysis showed the N4 sequence to group with the HRV-A consensus sequence in this region (17), this observation was misinterpreted. As demonstrated by phylogenetic analysis of this and flanking regions, all HRV-C variants genetically characterized in this region to date show evidence for recombination (Fig. 6) in a pattern remarkably similar to that observed in the 5′ UTR (with the noted exceptions of N10 and N4). The 2A protein of rhinoviruses and enteroviruses is a chymotrypsin-like proteinase of 142 amino acids. The region between positions 3268 to 3525 (residues 47 to 133) identified as recombinant corresponds precisely to the boundaries of the previously C-terminal structural domain of 2A, being comprised

FIG. 6. Comparison of the phylogeny of HRV-A, -B, and -C sequences in the putative recombinant region between position 3268 and 3525 with those of 5′ and 3′ flanking regions (positions 3009 to 3267 and 3526 to 3784). See the legend of Fig. 1 for tree construction method, labels, and symbols. For clarity, the accession numbers of species A and B variants have been prefixed by the species letter and serotype designation (e.g., A8 FJ445113 and B14 NC001490).

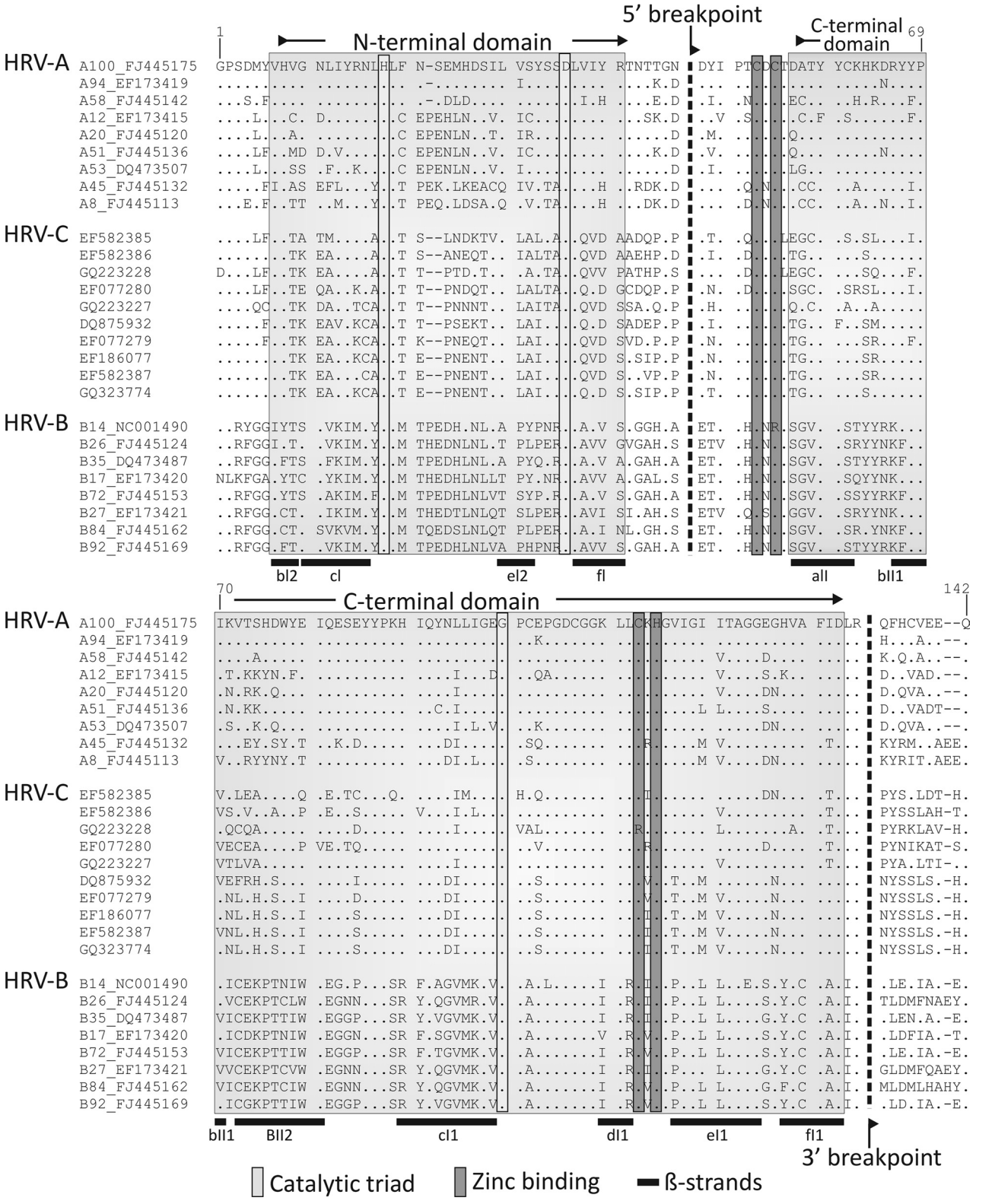


FIG. 7. Annotation of the secondary structure elements in the 2A gene of HRV-2 (40) superimposed on an alignment of the most genetically divergent HRV-A ($n = 9$) and HRV-B ($n = 8$) sequences and all available HRV-C 2A sequences and the inferred positions of the recombination breakpoints (dotted lines; nucleotide position 3268 to 3525). β-Strands and catalytic and zinc binding residues are as previously proposed (40). The positions of the N-terminal and C-terminal structural domains are indicated in shaded boxes. Numbering of amino acid residues follows that of the HRV-2 sequence X02316. For clarity, the accession numbers of species A and B variants have been prefixed by the species letter and serotype designation (as described in the legend of Fig. 6).

of six β-sheets that coordinate a zinc ion adjacent to the catalytic site of the protease (Fig. 7) (40). Our observations of recombination in the HRV-C 2A gene suggest at least some degree of modularity of its two structural domains. Investigation of the enzymatic properties of the HRV-C chimeric protein and the influence of the more variable N terminus of 2A encoded by HRV-C sequences on its function and specificity may provide insights into potential biological differences between HRV species. Finally, the proposed recombination junction at position 3268 coincides almost exactly with the 3′ end of the species A CRE (*cis*-acting RNA element) region (mapped to position 3226 to 3270 [13]). The absence of a homologous structure in the N′ terminal domain of species C variants is consistent with current predictions for its localization in VP2 (11).

The evolutionary events underlying the inferred 5′ UTR and 2A recombination events remain unclear. The existence of multiple recombination breakpoints in the 5′ UTR and the scattering of HRV-Ca genetic groups in the VP4/VP2 tree (17, 59) imply that recombination in the 5′ UTR has occurred several times independently. Similarly, the existence of multiple groupings of HRV-C sequences in the 2A clade of HRV-A suggests that more than one recombination event of that specific fragment occurred. What is remarkable is the evidence that recombination in these two regions is at least partly linked, with the same grouping of five of the HRV-C variants observed in both genomic regions. We are currently sequencing the 89 study samples in the 2A region to obtain additional information on the occurrence and linkage of recombination at these two sites in order to obtain a better understanding of the process and constraints under which these interspecies recombination events occur.

Genetic diversity of HRV-C. HRV-C variants showed substantially greater diversity than HRV-A or HRV-B serotypes (Fig. 3) but showed similar clustering of genetic groups into a number of well-defined (bootstrap-supported) clades (Fig. 1C) to other rhinovirus species. There were additionally two well-separated distributions of pairwise distances in VP4/VP2 and VP1 regions (Fig. 2B), suggesting that genetically defined types of HRV-C might be readily defined and demarcated for classification purposes. This genetic approach may be necessitated by the absence of HRV-C isolates with which to investigate serological interrelationships, as has been performed previously for HRV-A and -B species.

For other HRV species, there exists a small number of pairs of variants with genetic distances at an intermediate position in the distributions (the gray zone marked in Fig. 2 for HRV-A). For both HRV-A and -B, while there is a good correlation between VP1 sequence divergence and serological relationships (22, 26, 45), the latter are of limited value in defining a precise nucleotide sequence divergence threshold to separate inter- from intraserotype divergence values (22, 26). These discrepancies likely reflect the somewhat variable relationship between sequence divergence, epitope exposure, and antigenicity and the possibility that much of the sequence divergence between serotypes is immunologically driven.

For species C, however, the current data set of VP1 (and VP4/VP2) sequences shows a marked absence of variants showing intermediate distances in this sequence divergence range, and all can be defensibly assigned into different genetic groups using the 0.13 and 0.10 upper divergence thresholds for the VP1 and VP4/VP2 regions and into the same group if below. Also potentially facilitating a possible future genetically based classification of HRV-C is the lack of observed recombination between the VP4/VP2 and VP1 regions (Fig. 1B and C). This suggests that, once classified, HRV-C types may be readily identified using either VP4/VP2 sequences or those from VP1. This would be particularly helpful, given the preponderance of VP4/VP2 sequence data and the ease of amplification of this region for typing purposes. Remarkably, within 5 years of discovery of HRV-C, currently available sequences would correspond to a total of 60 types if classified in this way (Table 1).

Future classification proposals and the development of robust, well-defined type assignment criteria for HRV-C require discussion and consensus from an expert group, likely affiliated to the Picornavirus Study Group of the International Committee for the Taxonomy of Viruses. The data generated in the current study will be of value in future formulation of divergence thresholds in different genome regions for type assignment if this is to be adopted for classification purposes, and a consensus paper with formal proposals for type assignment based on the genetic groups defined in Table 1 has now been published (52a). Information on recombination frequency will similarly assist the interpretation of sequence data from other genomic regions and substantiate the proposed classification of Ca and Cc 5′ UTR variants (17) that we have extended in the current study. Other issues specific to HRV-C and which require consideration in formulating classification proposals include the lack of virus isolates or type strains equivalent to those of other rhinoviruses that currently precludes their serological and genetic/biological characterization.

ACKNOWLEDGMENTS

We are grateful to Elly Gaunt, Heli Harvala, Kate Templeton, Peter McCullough, Julie White, Mary Notman, Eleanor Leslie, and Carol Thomson for providing samples, data, and other virus testing results from the respiratory sample archive. We thank David J. Evans for valuable discussion of the manuscript.

Chloe McIntyre's Ph.D. studentship was funded by the Medical Research Council.

REFERENCES

1. **Anderson, E. C., S. L. Hunt, and R. J. Jackson.** 2007. Internal initiation of translation from the human rhinovirus-2 internal ribosome entry site requires the binding of Unr to two distinct sites on the 5′ untranslated region. J. Gen. Virol. **88:**3043–3052.
2. **Arden, K. E., C. E. Faux, N. T. O'Neill, P. McErlean, A. Nitsche, S. B. Lambert, M. D. Nissen, T. P. Sloots, and I. M. Mackay.** 2010. Molecular characterization and distinguishing features of a novel human rhinovirus (HRV) C, HRVC-QCE, detected in children with fever, cough and wheeze during 2003. J. Clin. Virol. **47:**219–223.
3. **Arden, K. E., P. McErlean, M. D. Nissen, T. P. Sloots, and I. M. Mackay.** 2006. Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections. J. Med. Virol. **78:**1232–1240.
4. **Benschop, K. S., C. H. Williams, K. C. Wolthers, G. Stanway, and P. Simmonds.** 2008. Widespread recombination within human parechoviruses: analysis of temporal dynamics and constraints. J. Gen. Virol. **89:**1030–1035.
5. **Blinkova, O., A. Kapoor, J. Victoria, M. Jones, N. Wolfe, A. Naeem, S. Shaukat, S. Sharif, M. M. Alam, M. Angez, S. Zaidi, and E. L. Delwart.** 2009. Cardioviruses are genetically diverse and cause common enteric infections in South Asian children. J. Virol. **83:**4631–4641.
6. **Borman, A., and R. J. Jackson.** 1992. Initiation of translation of human rhinovirus RNA: mapping the internal ribosome entry site. Virology **188:** 685–696.
7. **Briese, T., N. Renwick, M. Venter, R. G. Jarman, D. Ghosh, S. Kondgen,**

S. K. Shrestha, A. M. Hoegh, I. Casas, E. V. Adjogoua, C. Akoua-Koffi, K. S. Myint, D. T. Williams, G. Chidlow, B. R. van den, C. Calvo, O. Koch, G. Palacios, V. Kapoor, J. Villari, S. R. Dominguez, K. V. Holmes, G. Harnett, D. Smith, J. S. Mackenzie, H. Ellerbrok, B. Schweiger, K. Schonning, M. S. Chadha, F. H. Leendertz, A. C. Mishra, R. V. Gibbons, E. C. Holmes, and W. I. Lipkin. 2008. Global distribution of novel rhinovirus genotype. Emerg. Infect. Dis. **14:**944–947.

8. **Calvert, J., T. Chieochansin, K. Benschop, E. C. McWilliam-Leitch, J. F. Drexler, K. Grywna, H. da Costa Ribeiro, C. Drosten, H. Harvala, Y. Poovorawan, K. Wolthers, and P. Simmonds.** 2010. The recombination dynamics of human parechoviruses; investigation of type-specific differences in frequency and epidemiological correlates. J. Gen. Virol. **91:**1229–1238.
9. **Calvo, C., M. Luz Garcia, F. Pozo, N. Reyes, P. Perez-Brena, and I. Casas.** 2009. Role of rhinovirus C in apparently life-threatening events in infants, Spain. Emerg. Infect. Dis. **15:**1506–1508.
10. **Chevaliez, S., A. Szendroi, V. Caro, J. Balanant, S. Guillot, G. Berencsi, and F. Delpeyroux.** 2004. Molecular comparison of echovirus 11 strains circulating in Europe during an epidemic of multisystem hemorrhagic disease of infants indicates that evolution generally occurs by recombination. Virology **325:**56–70.
11. **Cordey, S., D. Gerlach, T. Junier, E. M. Zdobnov, L. Kaiser, and C. Tapparel.** 2008. The cis-acting replication elements define human enterovirus and rhinovirus species. RNA **14:**1568–1578.

11a. **Dominguez, S. R., T. Briese, G. Palacious, J. Hui, J. Villari, V. Kapoor, R. Tokarz, M. P. Glodé, M. S. Anderson, C. C. Robinson, K. V. Homes, and W. I. Lipkin.** 2008. Multiplex MassTag PCR for respiratory pathogens in pediatric nasopharyngeal washes negative by conventional diagnostic testing shows a high prevalence of viruses belonging to a newly recognized picornavirus clade. J. Clin. Virol. **43:**219–222.

12. **Drexler, J. F., S. Baumgarte, L. K. Souza Luna, A. Stocker, P. S. Almeida, T. C. Ribeiro, N. Petersen, P. Herzog, C. Pedroso, C. Brites, R. H. da Costa, Jr., A. Gmyl, C. Drosten, and A. Lukashev.** 2010. Genomic features and evolutionary constraints in Saffold-like Cardioviruses. J. Gen. Virol. **91:**1418–1427.
13. **Gerber, K., E. Wimmer, and A. V. Paul.** 2001. Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a *cis*-replicating element in the coding sequence of $2A^{pro}$. J. Virol. **75:**10979–10990.
14. **Hamparian, V. V., R. J. Colonno, M. K. Cooney, E. C. Dick, J. M. Gwaltney, Jr., J. H. Hughes, W. S. Jordan, Jr., A. Z. Kapikian, W. J. Mogabgab, A. Monto, C. A. Philips, R. R. Rueckert, J. H. Schieble, E. J. Stott, and D. A. J. Tyrrell.** 1987. A collaborative report: rhinoviruses—extension of the numbering system from 89 to 100. Virology **159:**191–192.
15. **Han, T. H., J. Y. Chung, E. S. Hwang, and J. W. Koo.** 2009. Detection of human rhinovirus C in children with acute lower respiratory tract infections in South Korea. Arch. Virol. **154:**987–991.
16. **Heath, L., W. E. van der, A. Varsani, and D. P. Martin.** 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J. Virol. **80:**11827–11832.
17. **Huang, T., W. Wang, M. Bessaud, P. Ren, J. Sheng, H. Yan, J. Zhang, X. Lin, Y. Wang, F. Delpeyroux, and V. Deubel.** 2009. Evidence of recombination and genetic diversity in human rhinoviruses in children with acute respiratory infection. PLoS One **4:**e6355.
18. **Kapikian, A. Z., R. M. Conant, V. V. Hamparian, R. M. Chanock, P. J. Chapple, E. C. Dick, J. D. Fenters, J. M. Gwaltney, Jr., D. Hamre, J. C. Holper, W. S. Jordan, E. H. Lennette, J. L. Melnick, W. J. Mogabgab, M. A. Mufson, C. A. Phillips, J. H. Schieble, and D. A. J. Tyrell.** 1967. Rhinoviruses: a numbering system. Nature **213:**761–762.
19. **Kapikian, A. Z., R. M. Conant, V. V. Hamparian, R. M. Chanock, E. C. Dick, J. M. Gwaltney, Jr., D. Hamre, W. S. Jordan, G. E. Kenny, E. H. Lennette, J. L. Melnick, W. J. Mogabgab, C. A. Phillips, J. H. Schieble, E. J. Stott, and D. A. J. Tyrell.** 1971. A collaborative report: rhinoviruses—extension of the numbering system. Virology **43:**524–526.
20. **Khetsuriani, N., X. Lu, W. G. Teague, N. Kazerouni, L. J. Anderson, and D. D. Erdman.** 2008. Novel human rhinoviruses and exacerbation of asthma in children. Emerg. Infect. Dis. **14:**1793–1796.
21. **Kistler, A., P. C. Avila, S. Rouskin, D. Wang, T. Ward, S. Yagi, D. Schnurr, D. Ganem, J. L. DeRisi, and H. A. Boushey.** 2007. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. J. Infect. Dis. **196:**817–825.
22. **Laine, P., C. Savolainen, S. Blomqvist, and T. Hovi.** 2005. Phylogenetic analysis of human rhinovirus capsid protein VP1 and 2A protease coding sequences confirms shared genus-like relationships with human enteroviruses. J. Gen. Virol. **86:**697–706.
23. **Lamson, D., N. Renwick, V. Kapoor, Z. Liu, G. Palacios, J. Ju, A. Dean, K. St George, T. Briese, and W. I. Lipkin.** 2006. MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004–2005. J. Infect. Dis. **194:**1398–1402.
24. **Lau, S. K., C. C. Yip, A. W. Lin, R. A. Lee, L. Y. So, Y. L. Lau, K. H. Chan, P. C. Woo, and K. Y. Yuen.** 2009. Clinical and molecular epidemiology of human rhinovirus C in children and adults in Hong Kong reveals a possible distinct human rhinovirus C subgroup. J. Infect. Dis. **200:**1096–1103.
25. **Lau, S. K., C. C. Yip, H. W. Tsoi, R. A. Lee, L. Y. So, Y. L. Lau, K. H. Chan, P. C. Woo, and K. Y. Yuen.** 2007. Clinical features and complete genome characterization of a distinct human rhinovirus (HRV) genetic cluster, probably representing a previously undetected HRV species, HRV-C, associated with acute respiratory illness in children. J. Clin. Microbiol. **45:**3655–3664.
26. **Ledford, R. M., N. R. Patel, T. M. Demenczuk, A. Watanyar, T. Herbertz, M. S. Collett, and D. C. Pevear.** 2004. VP1 sequencing of all human rhinovirus serotypes: insights into genus phylogeny and susceptibility to antiviral capsid-binding compounds. J. Virol. **78:**3663–3674.
27. **Lee, W. M., C. Kiesner, T. Pappas, I. Lee, K. Grindle, T. Jartti, B. Jakiela, R. F. Lemanske, Jr., P. A. Shult, and J. E. Gern.** 2007. A diverse group of previously unrecognized human rhinoviruses are common causes of respiratory illnesses in infants. PLoS One **2:**e966.
28. **Lindberg, A. M., P. Andersson, C. Savolainen, M. N. Mulders, and T. Hovi.** 2003. Evolution of the genome of human enterovirus B: incongruence between phylogenies of the VP1 and 3CD regions indicates frequent recombination within the species. J. Gen. Virol. **84:**1223–1235.
29. **Linsuwanon, P., S. Payungporn, R. Samransamruajkit, N. Posuwan, J. Makkoch, A. Theanboonlers, and Y. Poovorawan.** 2009. High prevalence of human rhinovirus C infection in Thai children with acute lower respiratory tract disease. J. Infect. **59:**115–121.
30. **Lukashev, A. N., V. A. Lashkevich, O. E. Ivanova, G. A. Koroleva, A. E. Hinkkanen, and J. Ilonen.** 2003. Recombination in circulating enteroviruses. J. Virol. **77:**10423–10431.
31. **Lukashev, A. N., V. A. Lashkevich, O. E. Ivanova, G. A. Koroleva, A. E. Hinkkanen, and J. Ilonen.** 2005. Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. J. Gen. Virol. **86:**3281–3290.
32. **McErlean, P., L. A. Shackelton, E. Andrews, D. R. Webster, S. B. Lambert, M. D. Nissen, T. P. Sloots, and I. M. Mackay.** 2008. Distinguishing molecular features and clinical characteristics of a putative new rhinovirus species, human rhinovirus C (HRV C). PLoS One **3:**e1847.
33. **McErlean, P., L. A. Shackelton, S. B. Lambert, M. D. Nissen, T. P. Sloots, and I. M. Mackay.** 2007. Characterisation of a newly identified human rhinovirus, HRV-QPM, discovered in infants with bronchiolitis. J. Clin. Virol. **39:**67–75.
34. **McKnight, K. L., and S. M. Lemon.** 1998. The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. RNA **4:**1569–1584.
35. **McWilliam Leitch, E. C., J. Bendig, M. Cabrerizo, J. Cardosa, T. Hyypia, O. E. Ivanova, A. Kelly, A. C. Kroes, A. Lukashev, A. Macadam, P. McMinn, M. Roivainen, G. Trallero, D. J. Evans, and P. Simmonds.** 2009. Transmission networks and population turnover of echovirus 30. J. Virol. **83:**2109–2118.
36. **Miller, E. K., K. M. Edwards, G. A. Weinberg, M. K. Iwane, M. R. Griffin, C. B. Hall, Y. Zhu, P. G. Szilagyi, L. L. Morin, L. H. Heil, X. Lu, and J. V. Williams.** 2009. A novel group of rhinoviruses is associated with asthma hospitalizations. J. Allergy Clin. Immunol. **123:**98–104.
37. **Miller, E. K., N. Khuri-Bulos, J. V. Williams, A. A. Shehabi, S. Faouri, J. Al, I., Q. Chen, L. Heil, Y. Mohamed, L. L. Morin, A. Ali, and N. B. Halasa.** 2009. Human rhinovirus C associated with wheezing in hospitalised children in the Middle East. J. Clin. Virol. **46:**85–89.
38. **Oprisan, G., M. Combiescu, S. Guillot, V. Caro, A. Combiescu, F. Delpeyroux, and R. Crainic.** 2002. Natural genetic recombination between co-circulating heterotypic enteroviruses. J. Gen. Virol. **83:**2193–2200.
39. **Palmenberg, A. C., D. Spiro, R. Kuzmickas, S. Wang, A. Djikeng, J. A. Rathe, C. M. Fraser-Liggett, and S. B. Liggett.** 2009. Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. Science **324:**55–59.
40. **Petersen, J. F., M. M. Cherney, H. D. Liebig, T. Skern, E. Kuechler, and M. N. James.** 1999. The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. EMBO J. **18:**5463–5475.
41. **Piralla, A., F. Rovida, G. Campanini, V. Rognoni, A. Marchi, F. Locatelli, and G. Gerna.** 2009. Clinical severity and molecular typing of human rhinovirus C strains during a fall outbreak affecting hospitalized patients. J. Clin. Virol. **45:**311–317.
42. **Renwick, N., B. Schweiger, V. Kapoor, Z. Liu, J. Villari, R. Bullmann, R. Miething, T. Briese, and W. I. Lipkin.** 2007. A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. J. Infect. Dis. **196:**1754–1760.
43. **Rivera, V. M., J. D. Welsh, and J. V. Maizel.** 1988. Comparative sequence analysis of the 5′ noncoding region of the enteroviruses and rhinoviruses. Virology **165:**42–50.
44. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
45. **Savolainen, C., S. Blomqvist, M. N. Mulders, and T. Hovi.** 2002. Genetic clustering of all 102 human rhinovirus prototype strains: serotype 87 is close to human enterovirus 70. J. Gen. Virol. **83:**333–340.
46. **Savolainen, C., P. Laine, M. N. Mulders, and T. Hovi.** 2004. Sequence

analysis of human rhinoviruses in the RNA-dependent RNA polymerase coding region reveals large within-species variation. J. Gen. Virol. **85:**2271–2277.

47. **Savolainen-Kopra, C., S. Blomqvist, S. Kaijalainen, U. Juonio, R. Juvonen, A. Peitso, A. Saukkoriipi, O. Vainio, T. Hovi, and M. Roivainen.** 2010. All known human rhinovirus species are present in sputum specimens of military recruits during respiratory infection. Viruses **1:**1178–1189.
48. **Savolainen-Kopra, C., S. Blomqvist, T. Kilpi, M. Roivainen, and T. Hovi.** 2009. Novel species of human rhinoviruses in acute otitis media. Pediatr. Infect. Dis. J. **28:**59–61.
49. **Savolainen-Kopra, C., S. Blomqvist, T. Smura, M. Roivainen, T. Hovi, D. Kiang, I. Kalra, S. Yagi, J. K. Louie, H. Boushey, J. Boothby, and D. P. Schnurr.** 2009. 5′ Noncoding region alone does not unequivocally determine genetic type of human rhinovirus strains. J. Clin. Microbiol. **47:**1278–1280.
50. **Simmonds, P.** 2006. Recombination and selection in the evolution of picornaviruses and other mammalian positive-stranded RNA viruses. J. Virol. **80:**11124–11140.
51. **Simmonds, P., and S. Midgley.** 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. **79:**15467–15476.
52. **Simmonds, P., and J. Welch.** 2006. Frequency and dynamics of recombination within different species of human enteroviruses. J. Virol. **80:**483–493.

52a. **Simmonds, P., C. L. McIntyre, C. Savolainen-Kopra, C. Tapparel, I. M. Mackay, and T. Hovi.** 21 July 2006. Proposals for the classification of human rhinovirus species C into genotypically-assigned types. J. Gen. Virol. doi: 10.1099/vir.0.023994-0.

53. **Stanway, G., F. Brown, P. Christian, T. Hovi, T. Hyypia, A. M. Q. King, N. J. Knowles, S. M. Lemon, P. D. Minor, M. A. Pallansch, A. C. Palmenberg, and T. Skern.** 2005. Family *Picornaviridae*, p. 757–778. *In* C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (ed.), Virus taxonomy: classification and nomenclature of viruses. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press, London, United Kingdom.
54. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
55. **Tamura, K., M. Nei, and S. Kumar.** 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U. S. A. **101:**11030–11035.
56. **Tapparel, C., T. Junier, D. Gerlach, S. Cordey, S. Van Belle, L. Perrin, E. M. Zdobnov, and L. Kaiser.** 2007. New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features. BMC Genomics **8:**224.
57. **Tapparel, C., T. Junier, D. Gerlach, S. Van Belle, L. Turin, S. Cordey, K. Muhlemann, N. Regamey, J. D. Aubert, P. M. Soccal, P. Eigenmann, E. Zdobnov, and L. Kaiser.** 2009. New respiratory enterovirus and recombinant rhinoviruses among circulating picornaviruses. Emerg. Infect. Dis. **15:**719–726.
58. **Tapparel, C., A. G. L'Huillier, A. L. Rougemont, M. Beghetti, C. Barazzone-Argiroffo, and L. Kaiser.** 2009. Pneumonia and pericarditis in a child with HRV-C infection: a case report. J. Clin. Virol. **45:**157–160.
59. **Wisdom, A., A. Kutkowska, E. C. McWilliam-Leitch, E. Gaunt, K. Templeton, H. Harvala, and P. Simmonds.** 2009. Genetics, recombination and clinical features of human rhinovirus species C (HRV-C) infections; interactions of HRV-C with other respiratory viruses. PLoS One **4:**e8518.
60. **Wisdom, A., C. McWilliam Leitch, E. Gaunt, H. Harvala, and P. Simmonds.** 2009. Screening respiratory samples for human rhinoviruses (HRV) and enteroviruses: comprehensive VP4/2-typing reveals high incidence and genetic diversity of HRV species C. J. Clin. Microbiol. **47:**3958–3967.

Arch Virol
DOI 10.1007/s00705-013-1634-6

ORIGINAL ARTICLE

Recombination in the evolution of human rhinovirus genomes

Chloe L. McIntyre · Carita Savolainen-Kopra · Tapani Hovi · Peter Simmonds

Received: 4 November 2012 / Accepted: 3 January 2013

Abstract Human rhinoviruses (HRV) are highly prevalent human respiratory pathogens that belong to the genus *Enterovirus*. Although recombination within the coding region is frequent in other picornavirus groups, most evidence of recombination in HRV has been restricted to the 5' untranslated region. We analysed the occurrence of recombination within published complete genome sequences of members of all three HRV species and additionally compared sequences from HRV strains spanning 14 years. HRV-B and HRV-C showed very little evidence of recombination within the coding region. In contrast, HRV-A sequences appeared to have undergone a large number of recombination events, typically involving whole type groups. This suggests that HRV-A may have been subject to extensive recombination during the period of diversification into types. This study demonstrates the rare and sporadic nature of contemporary recombination of HRV strains and contrasts with evidence of extensive recombination within HRV-A and between members of different species during earlier stages in its evolutionary diversification.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-013-1634-6) contains supplementary material, which is available to authorized users.

C. L. McIntyre (✉) · P. Simmonds
Infection and Immunity Division, Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, UK
e-mail: C.L.McIntyre@sms.ed.ac.uk

C. Savolainen-Kopra · T. Hovi
National Institute for Health and Welfare (THL), Mannerheimintie 166, FIN-00300, Helsinki, Finland

Introduction

Human rhinoviruses (HRVs) are a highly prevalent and genetically diverse group of human pathogens that cause a wide range of disease. Although they have been traditionally linked to mild, self-limiting upper respiratory tract infections, there is now evidence for a role of HRV in lower respiratory tract infection [13, 24, 31], exacerbations of chronic lung disease [20, 43, 70], severe systemic infections [66, 69] and gastrointestinal disease [14].

HRVs are single-stranded, positive-sense RNA viruses that belong to the genus *Enterovirus*, family *Picornaviridae*. HRVs are classified into three species on the basis of phylogenetic relationships and sequence divergence: *Human rhinovirus A*, *Human rhinovirus B*, and *Human rhinovirus C* [25, 49, 56]. While species A and B viruses can be readily isolated and cultured *in vitro* (and were among the earliest identified human picornaviruses), the equally prevalent HRV-C was only discovered very recently following the introduction of PCR-based diagnostic and surveillance screening methods [1, 21, 26, 28, 53]. For reasons that are currently poorly understood, HRV-C replicates poorly, if at all, in conventional cell culture, although recently, a variant of HRV-C was successfully cultured in sinus mucosal organ culture [4].

Serotypes of HRV-A and HRV-B were originally defined by serological cross-neutralisation assays [12, 19] and further characterised by receptor specificity and antiviral sensitivity profile. More recently, HRV-A and HRV-B have been genetically divided into two groups into which the 74 and 25 HRV types previously defined by serological methods have been assigned [56]. More recent analyses of phylogenetic relationships and genetic divergence among HRV strains has led to the reclassification of EV-68 as HRV-A87 [3] and of HRV-A21 and the Hanks strain as the

Published online: 27 February 2013

same genotype [27]. Three new HRV-A types have also been defined on the basis of sequence divergence in the VP1 region: HRV-A101–HRV-A103 (52, www.picornaviridae.com). For the more recently described species C rhinoviruses, difficulties associated with *in vitro* culture have precluded the development of serology-based virus typing, and variants are currently divided into 51 genotypically defined types. Classification into types is based on the current system used for enteroviruses [46], in this case based on an assignment threshold of 13 % nucleotide divergence in the VP1 region [59].

The epidemiological profile of HRVs and occurrence of infection with different types varies substantially every epidemic season [57], with high frequencies of asymptomatic infections [50] and no identified link between any strain, type or species of HRV with particular disease presentations. Despite an apparent difference in tissue tropism, the HRVs share many genetic similarities with other members of the genus *Enterovirus*. For instance, they have a single-stranded, positive-sense RNA genome of around 7100 bases with a 5' untranslated region (5'UTR) of around 600–650 bases. The coding region contains a single open reading frame and is cleaved post-translationally into four structural (VP1–VP4) and seven non-structural (2A–3D) proteins. The 5'UTR contains a type I internal ribosomal entry site (IRES) and a conserved 5' cloverleaf structure [48, 49].

EV serotypes, in particular EV-B, have been documented to undergo frequent recombination [30, 34, 40, 44, 61], and it has been proposed that structural and non-structural genome regions evolve independently [33, 45, 55]. EVs can thus be classified into a series of distinct recombinant forms (RFs), each possessing a phylogenetically distinct grouping in 3Dpol and other non-structural genes. These RFs have been shown to arise at intervals, dominate circulation for a few years and then disappear entirely [39, 41]. Recombination breakpoints in EV generally localise to two hotspot regions: between the 5'UTR and P1 and between P1 and P2 [61].

In contrast, recombination is generally thought to be rare and sporadic within the coding region of HRV [22, 29, 65], and a contemporary set of HRV-C sequences were found to cluster completely congruently in three coding genome regions [38]. To date, there is limited evidence of inter-species recombination within the coding region of HRV. It is well established, however, that HRV-A and HRV-C have undergone recombination within the 5'UTR, whereby over half of all HRV-C types group within the HRV-A clade [17, 68]. In addition, recent reports have highlighted the possibility that several contemporary HRV-A strains have been formed by recombination [49].

Until recently, data from the non-structural region of the HRV genome had only limited availability, as most HRV types had only one full genome sequence available. In addition, sequence data for HRV-C strains spanned a very limited time period due to their relatively recent discovery and the tendency of studies to focus on capsid-based typing. In the course of this analysis, we obtained sequences from isolates of all three HRV species spanning 14 years in order to determine whether recombination had shaped the contemporary diversity of these types. In addition, we also performed a comprehensive analysis of recombination between the VP1 and 3Dpol regions of all available HRV full genomes. This study represents an attempt to comprehensively characterise all detectable recombination events in the history of diversification of HRV.

Materials and methods

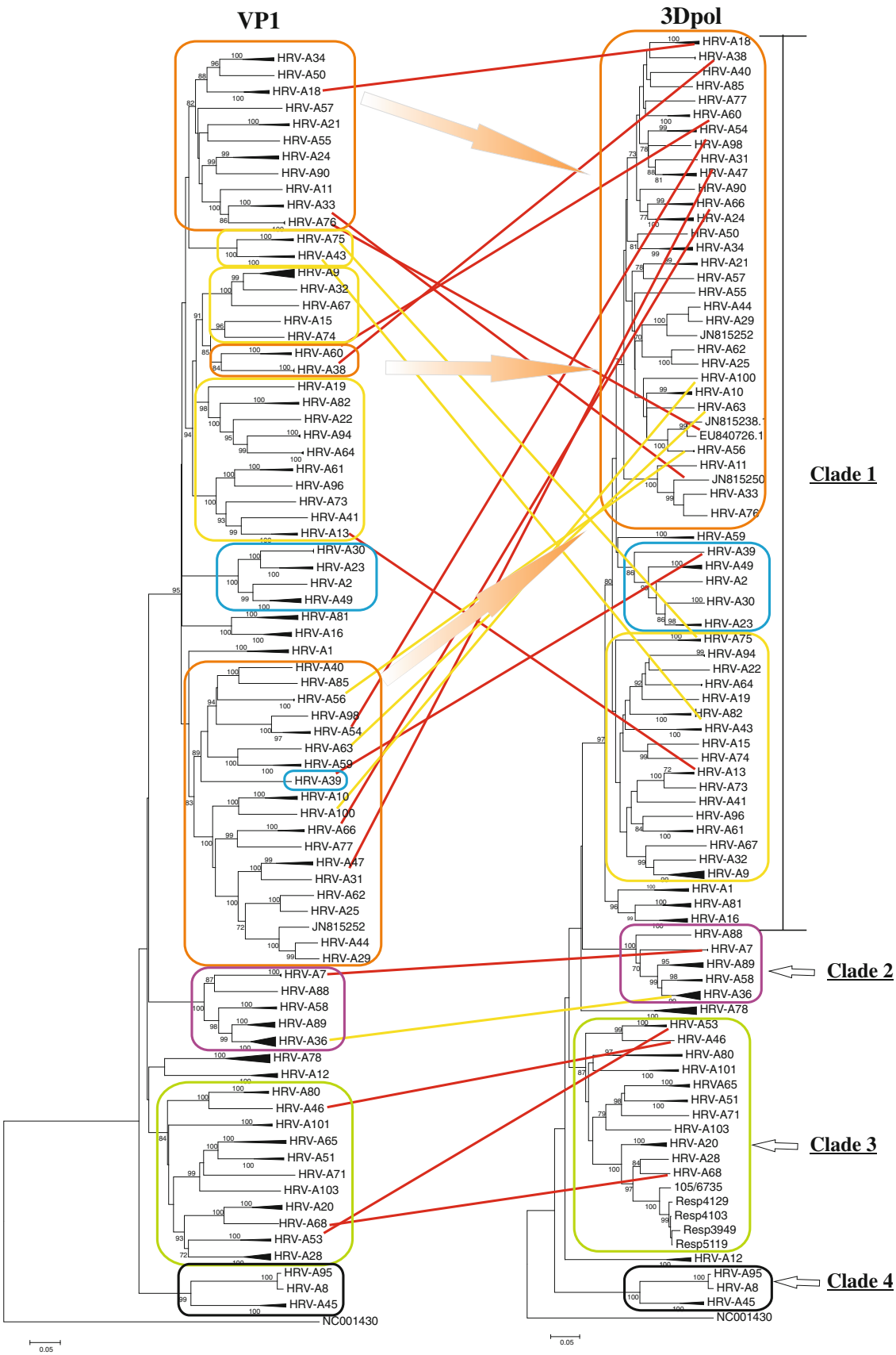
Sample selection

HRV-positive isolates were selected from distinct geographical locations (UK and Finland) over a period spanning 14 years. These included 28 HRV-positive clinical respiratory specimens from Finland collected between 1995 and 1997. VP4/VP2 sequences for these isolates were previously published with accession numbers in the series EU590043–EU590113 and AY015114–AY015174. Using previously defined pairwise nucleotide p-distance thresholds in the VP4/VP2 region [59, 67], these 28 samples were identified as belonging to 22 HRV types (10 HRV-A, 6 HRV-B and 6 HRV-C).

In addition, a total of 552 respiratory samples from patients previously referred to the Specialist Virology Centre at the Royal Infirmary of Edinburgh were screened by our previously described nested PCR method, specific for the 5'UTR region of HRV/HEV [67]. Positive samples (n=136) were amplified and sequenced in the VP4/VP2 region as described previously [67]. Samples that were of the same 22 types described above (n= 41) were identified and selected for further analysis. A further 23 HRV-C-positive samples were also identified, which represented 11 of the previously 28 provisionally assigned HRV-C types [59]. In total, 92 HRV isolates were analysed during the course of this work.

Amplification of the VP4/VP2, VP1 and 3Dpol regions of HRV-A, -B and –C

RNA was extracted from clinical samples using a QIAGEN Virus Spin Kit (QIAGEN, UK), according to manufacturer's instructions. Reverse transcription reactions were carried out using random hexamers and a Promega Access Reverse Transcription Kit with modifications as described previously [15]. Amplification of the VP1 and 3Dpol


VP1
3Dpol
HRV-A34
HRV-A50
HRV-A18
HRV-A57
HRV-A21
HRV-A55
HRV-A24
HRV-A90
HRV-A11
HRV-A33
HRV-A76
HRV-A75
HRV-A43
HRV-A9
HRV-A32
HRV-A67
HRV-A15
HRV-A74
HRV-A60
HRV-A38
HRV-A19
HRV-A82
HRV-A22
HRV-A94
HRV-A64
HRV-A61
HRV-A96
HRV-A73
HRV-A41
HRV-A13
HRV-A30
HRV-A23
HRV-A2
HRV-A49
HRV-A81
HRV-A16
HRV-A1
HRV-A40
HRV-A85
HRV-A56
HRV-A98
HRV-A54
HRV-A63
HRV-A59
HRV-A39
HRV-A10
HRV-A100
HRV-A66
HRV-A77
HRV-A47
HRV-A31
HRV-A62
HRV-A25
JN815252
HRV-A44
HRV-A29
HRV-A7
HRV-A88
HRV-A58
HRV-A89
HRV-A36
HRV-A78
HRV-A12
HRV-A80
HRV-A46
HRV-A101
HRV-A65
HRV-A51
HRV-A71
HRV-A103
HRV-A20
HRV-A68
HRV-A53
HRV-A28
HRV-A95
HRV-A8
HRV-A45
NC001430
0.05
HRV-A18
HRV-A38
HRV-A40
HRV-A85
HRV-A77
HRV-A60
HRV-A54
HRV-A98
HRV-A31
HRV-A47
HRV-A90
HRV-A66
HRV-A24
HRV-A50
HRV-A34
HRV-A21
HRV-A57
HRV-A55
HRV-A44
HRV-A29
JN815252
HRV-A62
HRV-A25
HRV-A100
HRV-A10
HRV-A63
JN815238.1
EU840726.1
HRV-A56
HRV-A11
JN815250
HRV-A33
HRV-A76
HRV-A59
HRV-A39
HRV-A49
HRV-A2
HRV-A30
HRV-A23
HRV-A75
HRV-A94
HRV-A22
HRV-A64
HRV-A19
HRV-A82
HRV-A43
HRV-A15
HRV-A74
HRV-A13
HRV-A73
HRV-A41
HRV-A96
HRV-A61
HRV-A67
HRV-A32
HRV-A9
HRV-A1
HRV-A81
HRV-A16
HRV-A88
HRV-A7
HRV-A89
HRV-A58
HRV-A36
HRV-A78
HRV-A53
HRV-A46
HRV-A80
HRV-A101
HRVA65
HRV-A51
HRV-A71
HRV-A103
HRV-A20
HRV-A28
HRV-A68
105/6735
Resp4129
Resp4103
Resp3949
Resp5119
HRV-A12
HRV-A95
HRV-A8
HRV-A45
NC001430
0.05
Clade 1
Clade 2
Clade 3
Clade 4

◀ **Fig. 1** Neighbour-joining phylogenetic trees showing the VP1 and partial 3Dpol regions of HRV-A strains. Where possible, HRV-A type groups have been collapsed for ease of reference. Contiguous clades are marked by coloured boxes. Clades 2, 3 and 4 are marked by purple, green and black boxes respectively, while clade 1 is divided into sub-clusters, which group together in 3Dpol. HRV-A types that undergo a bootstrap-supported change in nearest neighbour between VP1 and 3Dpol are marked with a red line. Changes in tree topology that are bootstrap supported on P1, P2 and P3 trees only are marked with a yellow line. Orange arrows indicate the formation of one 3Dpol clade from three VP1 clades

regions was carried out using the primers listed (Supplementary Table 1; Supplementary Data) and under the conditions described below.

HRV-A VP1

The full HRV-A VP1 region was assembled by amplification and sequencing of two overlapping sequence fragments. The same first-round product was utilised for second-round reactions for both VP1 fragments. For fragment 2, VP1 amplification was carried out as described previously [67]. For those samples that proved problematic to amplify, a combined reverse transcription and first-round PCR utilising the SuperScript III system (Invitrogen, UK) was carried out as described previously [42]. For fragment 1, the second round was a hemi-nested PCR reaction with primers 1943s and 2504as and previously described temperature cycling conditions. For HRV-A28-positive specimens, amplification of fragment 1 was not possible with standard primers, and a specific inner antisense primer was designed.

HRV-B VP1

Amplification of HRV-B VP1 was carried out using hemi-nested primers. PCR conditions were as for HRV-A, except for the use of an annealing temperature of 48 °C in the second round. For any samples that could not be amplified by this method, the SuperScript III system was used.

HRV-C VP1

SuperScript III was used to perform a combined reverse transcription and first-round PCR reaction directly on extracted RNA. The second round of the nested PCR reaction was carried out with an annealing temperature of 45 °C and in a 50-µl reaction volume. PCR products were loaded into wells on a 2 % agarose gel and separated at 150 V for 45 minutes. DNA bands of around 1000 bases were excised manually under UV transillumination and purified from the agarose gel using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, UK) according to the manufacturer's instructions.

HRV-A, -B and –C 3Dpol

Amplification of the partial 3Dpol region was carried out under previously described conditions [38] but using an annealing temperature of 45 °C in the second round. The amplified segment covered positions 6414–6896 in HRV-A (numbered according to FJ445111–HRV-A1), 6475-6960 in HRV-B (numbered according to X01087–HRV-B14) and 6361-6835 in HRV-C (numbered according to EF582385–HRV-C4).

All amplicons were sequenced using a BigDye Terminator Kit (Applied Biosystems, Warrington, UK). Sequences generated in this study have been submitted to GenBank and assigned accession numbers in the series KC342054–KC342173.

Sequence alignment and pairwise nucleotide p-distance calculations

All available HRV full genomes were downloaded from GenBank on 6/6/12. As a number of currently published HRV polyprotein sequences contain large gaps within the coding region, only those that were more than 90 % complete in the VP4/VP2, VP1 and partial 3Dpol regions were included. The regions analysed were 616–1002 for VP4/VP2, 2305–3126 for VP1 and 6361–6835 for partial 3Dpol (with all numbering according to reference strain HRV-C4, strain 024 (EF582385)). These comprised 136 HRV-A, 53 HRV-B and 121 HRV-C sequences, which were then compiled into separate datasets for each species, together with sequences generated in the current study: 28 HRV-A, 17 HRV-B and 24 HRV-C. A second dataset containing all available full genomes that were more than 90 % complete in the P1, P2 and P3 coding regions was constructed. Sequence alignments were performed in SSE v1.0 [58], using a combination of CLUSTAL/MUSCLE algorithms implemented in the program and manual inspection of nucleotide and amino acid alignments.

Pairwise nucleotide p-distances for individual regions were calculated using the program SequenceDist, within the SSE v1.0 package. For HRV-C, pairwise nucleotide p-distances in the VP1 region were used to confirm the minimum genetic divergence of the new HRV-C types [59].

Analysis of mean substitution rates

Sequences representing six HRV-A and HRV-C types were selected for analysis of mean substitution rates in the VP1 and partial 3Dpol regions. For VP1 datasets, every available GenBank entry that was >90 % complete across the region and had a date of isolation specified was included. 3Dpol datasets contained only sequences generated in this

study and published full genomes. HRV-A and HRV-C datasets were screened with GARD, SBP and RDP for evidence for intra- and inter-type recombination prior to substitution rate analysis.

Analysis of mean substitution rates was carried out in the Bayesian Evolutionary Analysis of Sampling Trees program (BEAST) [8]. Each analysis used the SRD06 model of nucleotide substitution, as recommended for protein coding data [8]. The analysis was carried out using a relaxed uncorrelated lognormal molecular clock and specified a constant population size, with auto-optimization of all other priors during execution of the program. Each MCMC chain was run for 20 million states, and the output was recorded every 1000 states. All datasets were analysed in duplicate to ensure convergence of results. In order to confirm that all observed results were a consequence of input sequence data, each dataset was subjected to an "empty" execution of BEAST with sequence data excluded. Additionally, all datasets were analysed with the dates of isolation of individual sequences scrambled.

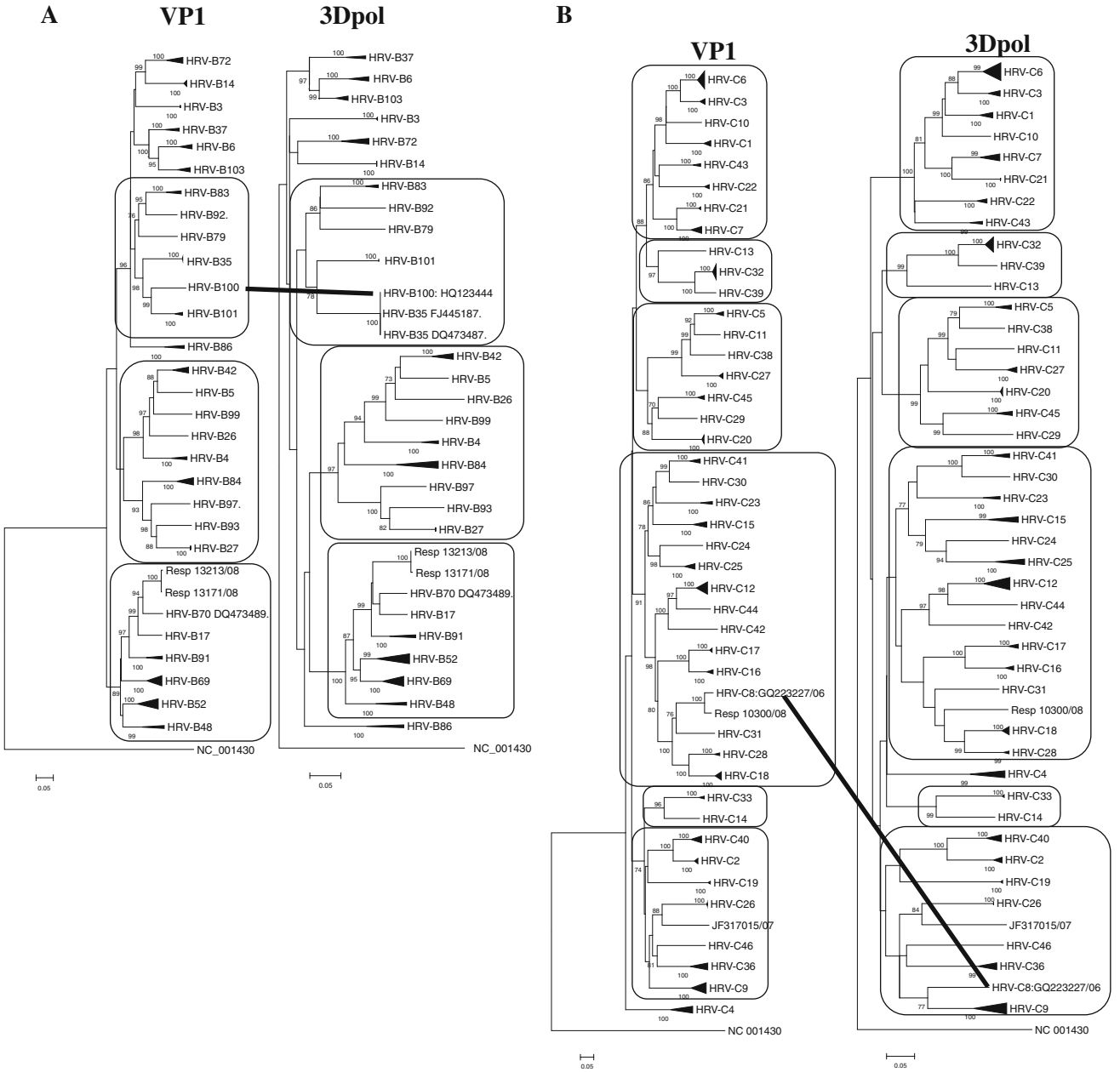


Fig. 2 **A**: Neighbour-joining phylogenetic trees of HRV-B VP1 and partial 3Dpol sequences. Putative recombinant sequences and type groups are marked with black lines. HRV-B type groups are collapsed where possible for ease of reference. Contiguous clades are marked with black boxes. **B**: Neighbour-joining phylogenetic trees of HRV-C VP1 and partial 3Dpol sequences. HRV-C type groups have been collapsed where possible for ease of reference. Contiguous clades are marked with black boxes

Table 1 Pairwise nucleotide p-distances of newly proposed HRV types to the nearest neighbours

GenBank accession number	Nearest neighbour (HRV type)	VP1 pairwise nucleotide p-distance	Provisional HRV-C type	Confirmed HRV type
JN815252	HRV-A44	HRV-A44: 0.109	-	-
	HRV-A29	HRV-A29: 0.1044		
JF781500	HRV-B35	0.1982	-	HRV-B101
JF781501	HRV-B35	0.1901	-	HRV-B101
JX074052	HRV-B35	0.1915	-	HRV-B101
JX074053	HRV-B35	0.2267	-	HRV-B102
JF416311[a]	HRV-C2	0.2900	HRV-Cpat1	HRV-C36
JF416321	HRV-C26	0.2719	HRV-Cpat2	HRV-C37
JF416322	HRV-C5	0.2147	HRV-Cpat4	HRV-C38
JF416306	HRV-C32	0.1472	HRV-Cpat5	HRV-C39
JF416312	HRV-C2	0.1755	HRV-Cpat6	HRV-C40
JF416323	HRV-C30	0.1904	HRV-Cpat7	HRV-C41
JF416320	HRV-C12	0.2506	HRV-Cpat8	HRV-C42
JF416307	HRV-C22	0.2460	HRV-Cpat9	HRV-C43
JF416310	HRV-C12	0.2077	HRV-Cpat11	HRV-C44
JF416308	HRV-C29	0.2540	HRV-Cpat12	HRV-C45
JF416318	HRV-C26	0.2836	HRV-Cpat25	HRV-C46

a For ease of reference, where more than one example of a previously provisionally assigned HRV-C type was sequenced in this study, only the earliest isolated strain is shown

Output was analysed in the TRACER program, within the BEAST package. Phylogenetic trees generated by Bayesian methods were annotated in TreeAnnotator v1.6 and visualised in FigTree. These were then directly compared to trees produced by bootstrap re-sampled maximum composite likelihood neighbour-joining trees to ensure consistency of phylogenetic relationships inferred using the two methods.

Phylogenetic analysis

Phylogenetic trees were constructed using the MEGA 5.0 software package [63] by the neighbour-joining method [54] from 100 bootstrap-resampled sequence alignments using the maximum composite likelihood (MCL) [11] distances with pairwise deletion for missing data. Phylogenetic trees were examined for changes in branch length and tree topology between non-consecutive genome fragments (VP4/VP2, VP1 and 3Dpol) and between coding regions of HRV full genomes (P1, P2 and P3).

Recombination analysis

RDP v4.0 [36] analysis was undertaken to further investigate observed phylogenetic incongruities. Putative recombinant sequences were analysed with their observed nearest neighbours in each region. Each sequence set was examined using a combination of algorithms, including RDP [37], GeneConv [47], MaxChi [62], Chimaera [51], SiScan [10] and Bootscan [35]. A potential recombination event was accepted for further analysis if it was detected by more than two of the above algorithms, with a p-value of less than 0.05.

All datasets were further analysed using the GARD and SBP [23] programs available on the HyPhy datamonkey webserver (www.datamonkey.org). Prior to analysis of each dataset, the model selection tool included was employed, and the appropriate model of nucleotide substitution was used for further analysis.

In addition, as a control measure, several groups of sequences that were observed to display congruent phylogenetic relationships between VP1 and 3Dpol were analysed by these methods. All analyses of putative non-recombinant sequence groups confirmed that there was no evidence of recombination.

The combination of different recombination detection algorithms detailed above were used as a screening tool in order to effectively screen each alignment for evidence of recombination. If evidence of recombination was detected by these methods, the recorded event was further analysed by phylogenetic analysis of putative recombinant regions and further characterised by analysis with the GroupScan program within the SSE v1.0 software package.

Determining putative recombination breakpoints

In order to verify putative recombination events and determine recombination breakpoints where possible, sets

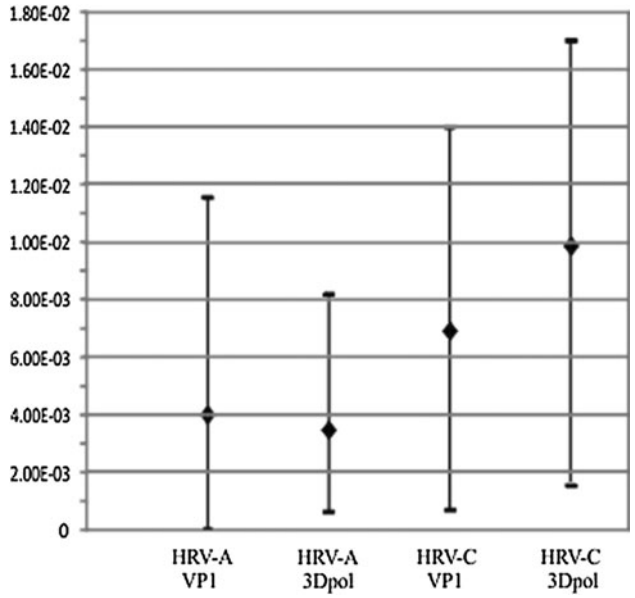


Fig. 3 Mean substitution rates and 95 % highest posterior density intervals for two coding regions of selected HRV-A and HRV-C sequences

of sequences underwent GroupScan analysis in the SSEv1.0 software package. GroupScan scores the extent of phylogenetic grouping of a query sequence with predefined sequence groups and can pinpoint potential breakpoints if sequences that are closely related to both parental groups are included. The GroupScan method relies on detection of bootstrap-supported changes in tree topology along an alignment and was used for the determination of recombination breakpoints. For each analysis, the three sequenced regions of each isolate were concatenated into a single sequence, and a fragment size of 300 bases with an increment of 30 bases was used. Additionally, each analysis was repeated with a fragment size of 600 bases and an increment of 30 to ensure consistency of results. In each instance, the query sequence was the potential recombinant sequence/group and was scanned against the nearest neighbour/group in both regions. The intersection of the two lines was calculated to give an estimation of the recombination breakpoint. Breakpoints were compared and found to be highly similar in scans carried out with a 300-nucleotide fragment and those done with a 600-nucleotide fragment.

Results

Genotype assignment for HRV-A, HRV-B and HRV-C sequences

A total of 164 HRV-A, 70 HRV-B and 145 HRV-C variants that were more than 90 % complete across the analysed regions were included in the recombination analysis. A separate recombination analysis was performed with complete genome sequences (201 HRV-A, 71 HRV-B and 59 HRV-C).

Examination of phylogenetic trees constructed for the VP1 region of members of all three HRV species revealed consistent grouping of individual (sero)types into clades with uniformly high bootstrap support (Figs. 1, 2). These type groupings were also apparent in the VP4/VP2 and full P1 capsid coding region. However, the relationships between HRV type groups were generally less well resolved in VP4/VP2 trees as a result of greater sequence conservation and shorter fragment length (data not shown).

Previously proposed pairwise nucleotide p-distance divergence thresholds were used to support phylogenetic genotype groupings and define putative new HRV types [59, 67] (Table 1).

One newly described HRV-A sequence, JN815252, did not group with any currently assigned HRV-A type and shared nearest neighbours, HRV-A29 (FJ445125) and HRV-A44 (DQ473499), in every region studied. Using the previously determined pairwise nucleotide p-distance threshold of 12 % divergence in VP1 for identification of new HRV-A types [67], JN815252 fell within the intra-serotype distance range on comparison with both HRV-A29 and HRV-A44 (0.1051 and 0.1083, respectively). Divergence between HRV-A29 and HRV-A44 has previously been shown to fall within the intra-serotype distance range [67], and these two serotypes have also been demonstrated to be serologically cross-reactive [7]. These three sequences were treated as the same serotype for recombination analysis.

In species B, JF781500, JF781501 and JX074052 formed a bootstrap-supported clade in VP1 with a pairwise nucleotide p-distance of 0.015 to 0.018 from each other and of consistently greater than 0.19 from the nearest assigned HRV-B type (HRV-B35). JX074053 also displayed a p-distance of 0.2267 from HRV-B35, its nearest relative. By current criteria [67], these two groups represent new HRV-B types and have been formally designated HRV-B101 and HRV-B102 (Table 1) (http://www.picornastudygroup.com/types/enterovirus/hrv-b.htm).

HRV-C VP1 sequence data generated during this study confirmed type assignment of eleven HRV-C types that had been provisionally identified as new types based on VP4/VP2 sequences [59]. These had pairwise distances that fell above the VP1 divergence threshold (Table 1) and have been formally designated as HRV-C36 – HRV-C46 (http://www.picornastudygroup.com/types/enterovirus/hrv-c.htm). The previously proposed nucleotide divergence threshold of 13 % was readily identifiable in the distribution of pairwise VP1 p-distances generated from the expanded VP1 dataset (Fig. S1, Supplementary Data).

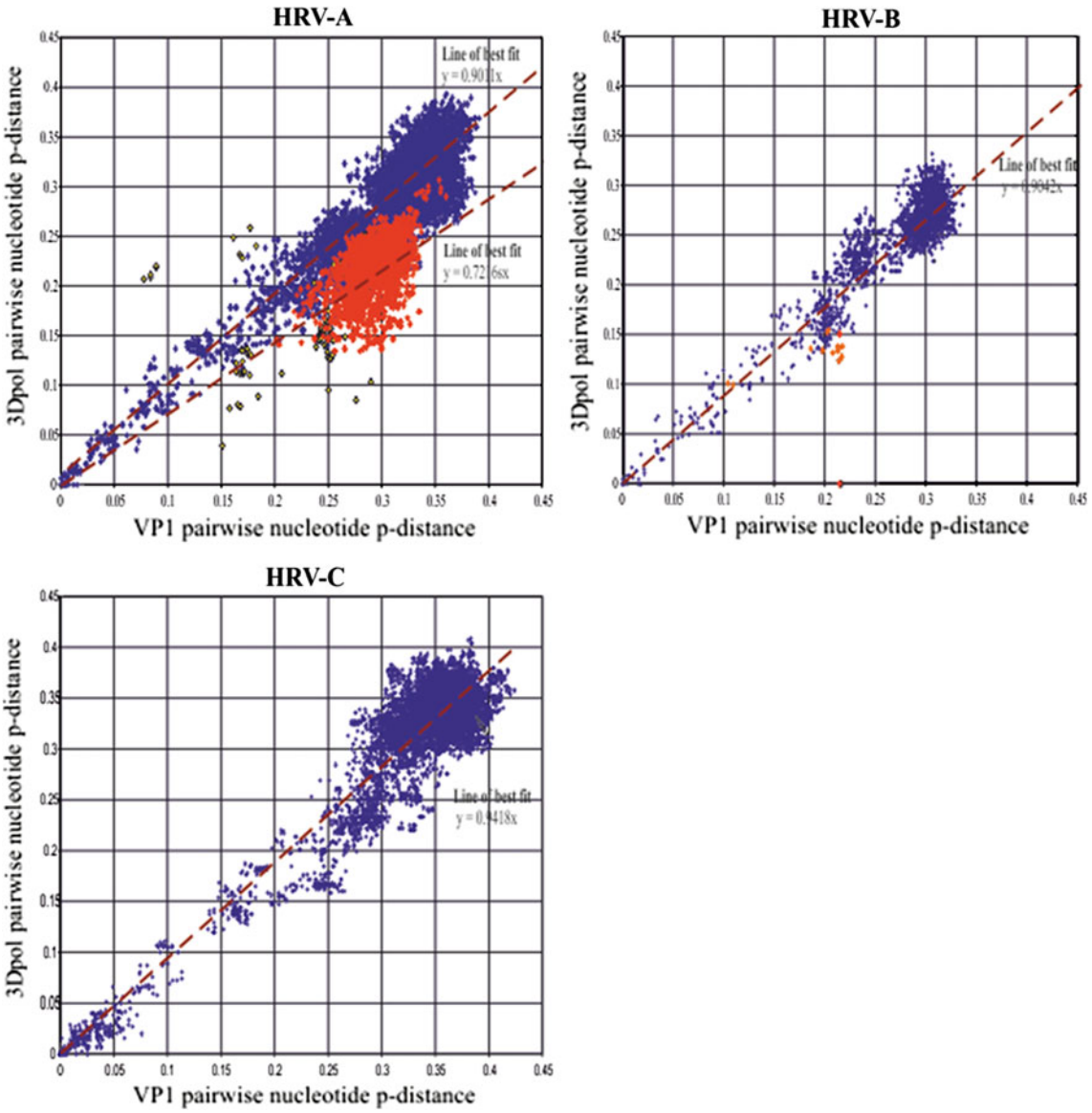


Fig. 4 Comparison of pairwise nucleotide p-distances between VP1 and partial 3Dpol regions of HRV-A, HRV-B and HRV-C. The line of best fit is indicated by a dotted red line. For HRV-A, the line of best fit is given separately for putative recombinant and non-recombinant distributions. A graph depicting HRV-A pairwise comparisons shows putative recombination events involving full HRV-A type groups [such as observed extensively within clade 1 (Fig. 4)] in red. Sporadic recombination events involving single HRV-A sequences are marked in yellow. For graphs depicting HRV-B and HRV-C, all examples of putative recombination events are marked in red

Analysis of mean substitution rates in selected HRV-A and HRV-C types

Three non-recombinant types each of HRV-A (HRV-A9, -A28 and –A78) and HRV-C (HRV-C9, -C12 and –C18) with dates of isolation spanning at least 13 years were selected for analysis of mean substitution rates. The HRV-A dataset assembled contained sequences spanning 40 years, while HRV-C sequences spanned 14 years. Phylogenetic trees analysed for each type included showed no bootstrap-supported change in topology within each type or species (data not shown). Insufficient sequences of individual HRV-B types were available for inclusion of members of this species in the analysis. Analysis was carried out separately for HRV-A sequences (n=18) and HRV-C sequences (n=23).

For both analysed regions of HRV-A and HRV-C, mean substitution rates and 95 % highest posterior density (HPD) intervals were on the order of 10^{-3}–10^{-2} substitutions per site per year. Calculated mean substitution rates were largely similar between the two coding regions and two HRV species considered (Fig. 3). Although both regions of HRV-C showed 2- to 3-fold elevated substitution rates (6.9138×10^{-3} and 9.8491×10^{-3} for VP1 and 3Dpol, respectively) when compared to the estimates obtained for

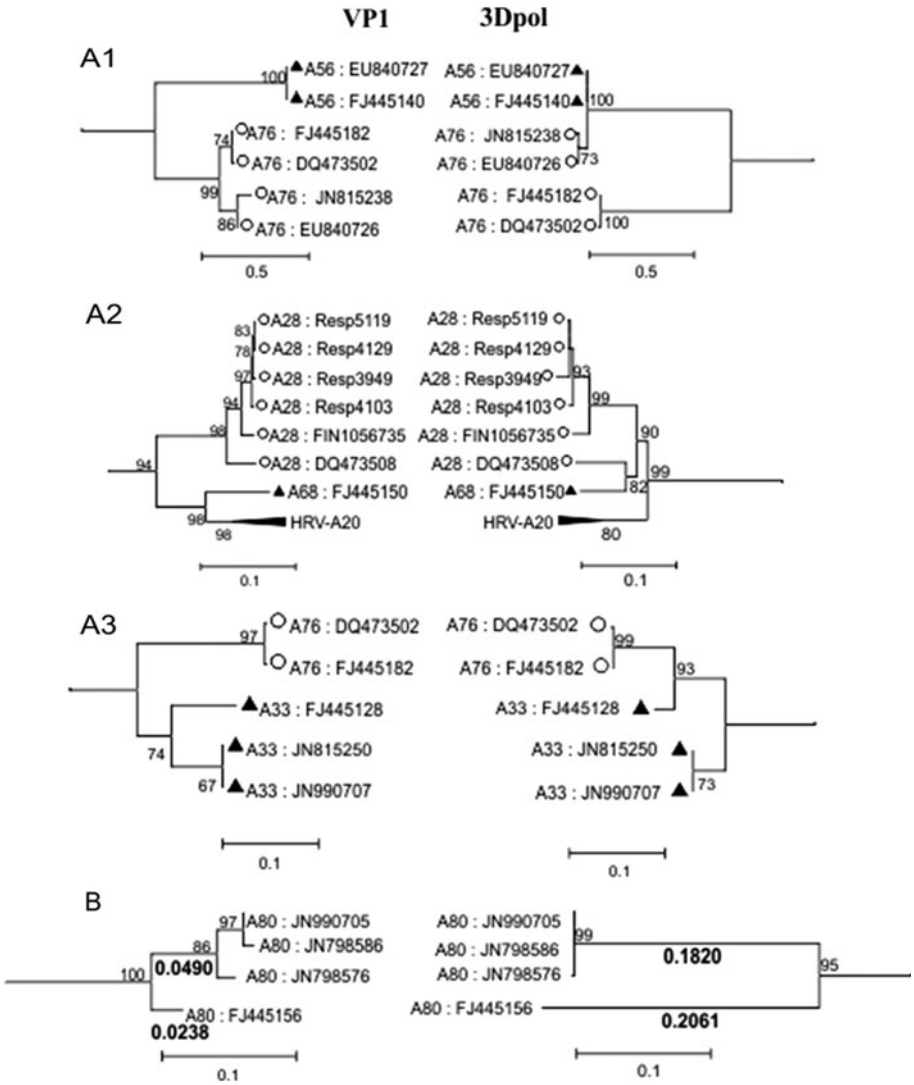


Fig. 5 Neighbour-joining phylogenetic trees representing distinct patterns of recombination observed within HRV-A sequences. The branch to the tree root has been collapsed, and the outgroup used for analysis is given below. **A**: Phylogenetic trees showing examples of incongruent topology between VP1 and 3Dpol involving [1] HRV-A56 and HRV-A76 (outgroup: A7), [2] HRV-A28 and HRV-A68 (outgroup: A95) and [3] HRV-A33 and older HRV-A76 strains (outgroup: A88) **B**: Phylogenetic tree showing branch-length discrepancies between VP1 and 3Dpol of HRV-A80 (outgroup: A46)

HRV-A strains (3.9954×10^{-3} and 3.4586×10^{-3} for VP1 and 3Dpol, respectively), the 95 % HPD intervals overlapped for the two species.

In order to ensure that the specified priors were not inadvertently influencing the posterior distribution of the calculated mean substitution rate, the BEAST algorithm was executed with sequence data excluded for each dataset. As additional confirmation, dates of isolation for individual sequences were randomised using the SSE v1.0 software package, and analysis with BEAST was repeated. Scrambled datasets gave extremely low substitution rate estimates well below the lower HPD interval of the test sequences (data not shown).

Comparison of pairwise nucleotide p-distances in VP1 and 3Dpol

The similarity of mean substitution rates between VP1 and 3Dpol regions in HRV-A and HRV-C is consistent with the observed linear relationships between the pairwise p-distances of sequences in these two regions (Fig. 4). This linear relationship was observed in HRV-B and HRV-C, in both intra- and inter-(sero) type distance ranges (Fig. 4). Both distributions contained few if any outlying data points, and lines of best fit had gradients of approximately 1 (0.9042 and 0.9418, respectively). This correlation provides evidence for maintained equal substitutions rates in the two regions throughout the period of diversification into different types.

In contrast, HRV-A sequences showed lower divergence in the 3Dpol region compared to VP1 and a gradient of 0.8714 for the line of best fit (Fig. 4). The distribution additionally contained a large number of outlying data points, consistent with the occurrence of several recombination events during HRV-A diversification. HRV-A pairwise distances could be divided into three distinct distributions, consisting of non-recombinant types, potential recombination events involving full HRV-A type

groups and sporadic instances of phylogenetic incongruity involving single HRV-A sequences (marked with blue, red and yellow, respectively, in Fig. 4; see phylogenetic analysis below). For the putative non-recombinant types that grouped consistently in VP1 and 3Dpol regions, a gradient of approximately 1 (0.9011) was observed. Meanwhile, consideration of pairwise p-distance comparisons between potentially recombinant HRV-A sequences alone gave a calculated gradient of 0.7216.

Analysis of phylogenetic incongruity in HRV-B and HRV-C sequences

In accordance with the observed conformity of HRV-B VP1 and 3Dpol pairwise distances (Fig. 4), the majority of HRV-B sequences displayed congruent clustering in phylogenetic trees and clustered within their genotype group, as defined by VP1 (Fig. 2A). The only bootstrap-supported putative recombinant sequence observed was HQ123444. This full genome sequence showed a pairwise p-distance of at least 0.19 from HRV-B35 strains in VP1. However, on inspection of all published HRV full genomes, HQ123444 was identical to HRV-B35 from position 6353 in the 3Dpol region. This incongruity was detected using RDP, GARD and SBP.

In addition, a closer phylogenetic grouping of the HRV-B17 and HRV-B70 type strains, as compared to contemporary HRV-B70 sequences, was observed in the 3Dpol region. Although this change of topology did not have adequate support in 3Dpol (Fig. 2A), it was supported in both the P2 and P3 regions (data not shown).

The majority of HRV-C sequences fell within six bootstrap-supported clades, which were congruent between all regions analysed (Fig. 2B). With the previously reported artefactually recombinant sequence GQ223227 excluded from analysis [17, 38], no evidence of inter-type recombination was observed in HRV-C. Any observed inconsistency in branching order or phylogenetic relationships between HRV-C types could not be confirmed by analysis with RDP, GARD or SBP.

Analysis of phylogenetic incongruity in HRV-A sequences

In contrast to HRV-B and HRV-C, the comparison of VP1 and partial 3Dpol pairwise nucleotide p-distances in HRV-A suggested a number of separate recombination events between VP1 and 3Dpol (Fig. 1) and between the P1, P2 and P3 regions (Fig. S2; Supplementary data). Putative recombination events identified by phylogenetic analysis were confirmed by analysis with RDP, GARD and SBP.

Several HRV-A types showed a pattern similar to that of HRV-B70, with contemporary strains grouping separately from their respective older type strains in the non-structural protein coding region. As reported previously [65], HRV-A76 strains collected between 1999 and 2010 grouped most closely with HRV-A56 in 3Dpol (Fig. 5 A1). The intra-type pairwise nucleotide p-distances of HRV-A76 strains ranged from 0.0839 to 0.0933 in VP1 and from 0.2085 to 0.2292 in 3Dpol. HRV-A68 and HRV-A33 also displayed a similar pattern, whereby the older strains grouped preferentially with HRV-A28 (Fig. 5 A2) and HRV-A76 (Fig. 5 A3) type strains, respectively, in 3Dpol. The HRV-A40 type group includes three recently described full genome sequences that were not included in VP1 and 3Dpol analysis due to gaps present in published sequences, which resulted in sequence completeness below the threshold of 90 %. However, on analysis of the whole P3 region, the three contemporary HRV-A40 strains (JX074051, JQ245067 and JN798579) formed a bootstrap-supported clade separate from the HRV-A40 type strain, which maintained its grouping with HRV-A85.

HRV-A80 sequences showed a substantial difference in branch lengths between two regions, with no accompanying change in tree topology (Fig. 5B). The HRV-A80 type strain had a pairwise nucleotide p-distance of 0.078 from contemporary HRV-A80 strains in VP1. However, in 3Dpol, the contemporary HRV-A80 strains were much more divergent from the type strain, with a p-distance of 0.207. All of the above putative recombination events were verified by analysis with GARD, SBP and RDP.

With the exception of HRV-A12 and HRV-A78, all HRV-A sequences grouped into four bootstrap-supported clades (numbered 1-4 in Fig. 1) branching basally in both VP1 and 3Dpol trees. Clades 1–4 were maintained in all regions studied, and although bootstrap-supported changes in tree topology were relatively common within each of these clades, there were no instances of recombination occurring between them.

Bootstrap-supported changes in branching order and tree topology were observed within clades 1, 2 and 3 (marked with red lines in Fig. 1). The number of bootstrap-supported nodes that each putative recombinant sequence violated in the transition between VP1 and 3Dpol grouping ranged from two to eight (Table 2). Clade 1 formed several sub-clusters of HRV-A types, and there was a clear breakdown in phylogenetic relationships of HRV-A types between VP1 and 3Dpol. For example, one monophyletic bootstrap-supported clade in 3Dpol (orange box; Fig. 1) was formed from three distinct bootstrap-supported VP1 clades. Phylogenetic relationships between involved HRV-A types within this 3Dpol clade were largely incongruent when compared with the VP1 tree topology. However, there was poor bootstrap support for nearest-neighbour relationships between HRV-A types in this region, regardless of whether 3Dpol or full P3 region trees were inspected. Notably, the large cluster of putative recombinant HRV-A pairwise nucleotide p-distance

Table 2 Phylogenetic incongruities observed between VP1 and 3Dpol in HRV-A sequences

HRV type	Accession numbers	Nearest neighbour					Number of changes[a]			B/p mean (range)
		VP1	3Dpol	P1	P2	P3	VP1 → 3Dpol	P1 → P2	P2 → P3	
A13	FJ445116	A41	A73	A41	A73	A73	2	2	0	3393
	FJ445117									(3365–3420)
A53	DQ473507	A28	A46	A28	A46	A46	4	4	0	3213
	JN798587									(3198–3228)
A46	DQ473506	A80	A53	A80	A53	A53	4	4	0	3280
A68	FJ445150	A20	A28	A20	A20	A28	4	0	3	3998
	JN798578[b]									(3995–4001)
A7	FJ445176	A88	A36	A88	A88	A88	3	0	0	6576
	DQ473503		A58		A36	A36				
			A89		A58	A58				
					A89	A89				
A76	EU840726	A76	A56	A76	A76	A56	7	0	7	5130
(except FJ445182	JN815238									(5129–5131)
DQ473502)	**JX074055**									
	JX074049									
A54	FJ445138	A56	A31	A56	A40	A31	6	2	5	3316
A98	FJ445139 FJ445173	A40	A47	A40	A85	A47				(3238–3377)
		A85		A85						
A31	FJ445126	A62	A98	A62	A100	A98	6	0	5	5224
A47	FJ445133	A25	A54	A25	A62	A54				5424
	JN837692	A44		A44	A25					(5416–5435)
	GQ223229	A29		A29	A44					
		JN815252		JN815252	A29					
					JN815252					
A18	FJ445118	A34	A38	A34	A66	A85	7[d]	9	5	3397
	F292/8643[c]	A50	A40	A50	A77					(3328–3526)
	JF781496		A85							5276
	JF781510		A77							(5024–5577)
			A60							
			A54							
			A98							
			A31							
			A47							
			A90							
			A66							
			A24							
A39	AY751783	A63	A2	A63	A2	A2	4	5	0	3431
		A59	A23	A59	A23	A23				
		A98	A30	A98	A30	A30				
		A54	A49	A54	A49	A49				
		A56		A56						
		A40		A40						
		A85		A85						
A66	FJ445148	A77	A24	A77	A77	A24	6	0	6	5652
	JN112340					A90				(5645–5674)
	JN621246									
	JQ837715									

Table 2 continued

HRV type	Accession numbers	Nearest neighbour					Number of changes[a]			B/p mean (range)
		VP1	3Dpol	P1	P2	P3	VP1 → 3Dpol	P1 → P2	P2 → P3	
A38	FJ445180	A15	A18	A15	A19	A18	7[e]	4	14[f]	5130
	DQ473495	A74	A40	A74	A60	A40				(5118–5143)
	JQ994496	A60	A85	A9		A85				
	JN541272		A77	A32		A77				
			A54	A67		A54				
			A98	A60		A98				
			A31			A31				
			A47			A47				
			A90			A60				
			A66							
			A24							
			A60							
A60	FJ445143	A15	A18	A15	A19	A18	8[g]	3[h]	10	5117
	JN798590	A74	A40	A74	A60	A40				(5111–5123)
		A60	A85	A9		A85				
			A77	A32		A77				
			A54	A67		A54				
			A98	A60		A98				
			A31			A31				
			A47			A47				
			A90			A60				
			A66							
			A24							
			A60							

a "Number of changes" refers to the number of bootstrap-supported nodes a group of sequences violates to form a new grouping

b Sequences indicated in bold font are those that are not present on VP1 and 3Dpol trees due to sequence gaps

c Sequences indicated in bold italics are those that consist only of non-consecutive sequence fragments (VP4/VP2, VP1 and 3Dpol) and are therefore not represented on P1, P2 and P3 trees

d Due to lack of bootstrap support for nearest-neighbour grouping in 3Dpol for A18, node changes are taken from grouping with HRV-A34 to HRV-A85 (nearest bootstrap supported neighbour in P3 region)

e Similarly, node changes for HRV-A38 between VP1 and 3Dpol are taken from grouping with HRV-A60 to HRV-A18

f Node changes for HRV-A38 between P2 and P3 are taken from HRV-A60 to HRV-A40

g Node changes for HRV-A60 between VP1 and 3Dpol are taken from HRV-A38 to HRV-A77

h Node changes for HRV-A60 between P2 and P3 are taken from HRV-A38 to HRV-A40

comparisons between 0.25 and 0.32 VP1 p-distance (coloured red in Fig. 4) corresponded to the formation of this new 3Dpol clade. Putative recombination events involving HRV-A types from all three clades were further analysed (Table 2).

Sequence groupings in phylogenetic trees constructed for whole P1, P2 and P3 coding regions were largely consistent with those observed in VP1 and 3Dpol; P1 was similar to the VP1 tree and P2/P2 was similar to 3Dpol (Fig. S2; Supplementary Data). Sequence analysis of these longer regions did, however, identify bootstrap-supported phylogenetic incongruity of a further six HRV-A types between regions (yellow lines in Fig. 1; listed in Table S1; Supplementary Data).

All phylogenetic incongruities listed were additionally confirmed by RDP, GARD and SBP analysis (data not shown).

Analysis of putative recombination breakpoints

In order to determine putative breakpoints in HRV-A sequences showing evidence for recombination between VP1 and 3Dpol, full genomes from each HRV-A type underwent GroupScan analysis and were compared directly to their nearest neighbours in P1, P2 and P3. Two representative examples of Grouping Scan output are given, with putative recombination breakpoints highlighted (Fig. 6A). HRV-A18 has a breakpoint near the P1/P2

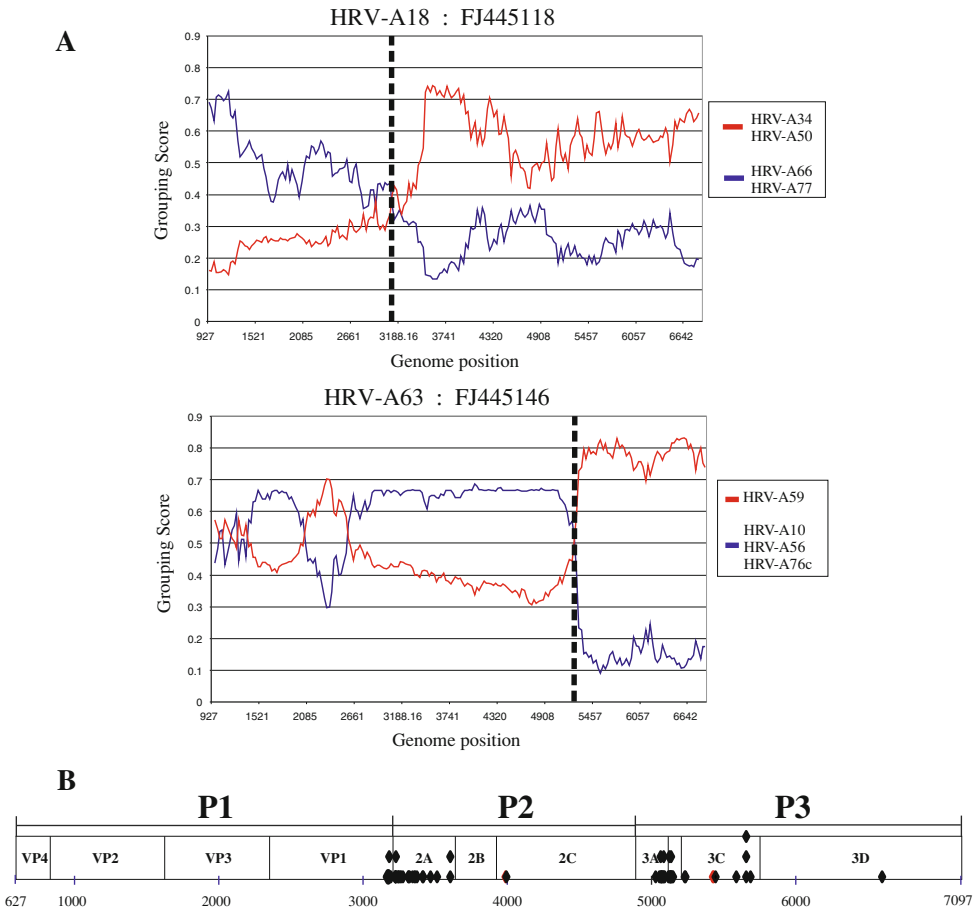


Fig. 6 **A**: GroupScan of two representative recombinant HRV-A types. The proposed recombination breakpoint is indicated by a dotted line. **B**: Diagram showing calculated coding region recombination breakpoints for all HRV-A recombinant types, mapped onto a diagram of the full genome of HRV-A sequences (numbering taken from FJ445111–HRV-A1). Those from the large recombinant 3Dpol clade, which lacked bootstrap support for nearest-neighbour groupings in the analysed regions, are excluded

boundary, and HRV-A63 has a breakpoint within P3. The genome fragment between positions 2000 and 2500 in HRV-A63, which shows a similarly elevated grouping score with HRV-A59, did not correspond to any bootstrap-supported change in tree topology in this region (data not shown).

The majority of breakpoints occur near the P1/P2 boundary, specifically within the 2A coding region (Fig. 6B). There was also a second putative hotspot region around the P2/P3 boundary, and most of these occurred within the 3A coding region or at the 3A/3B junction.

Recombination in the 5' untranslated region of HRV-A, -B and –C

All sequences that were more than 90 % complete across the 5'UTR fragment numbered 167–626 (numbered by FJ445111) were included in the analysis (167 HRV-A, 38 HRV-B and 59 HRV-C). The occurrence of recombination in the 5'UTR was assessed by visual inspection of phylogenetic trees for HRV-A (Fig. 7A), HRV-B (Fig. 7B) and HRV-A/HRV-C (Fig. S3: Supplementary Data).

All 5'UTR sequences from HRV-B clustered together and were distinct from those of HRV-A and HRV-C (data not shown). In contrast, HRV-A and over two-thirds (69 %) of HRV-C sequences were interspersed in the 5'UTR, consistent with inter-species recombination, as proposed previously [17, 38, 68].

Inspection of HRV-A sequences exclusively revealed 17 putative recombinant sequences, from nine HRV-A types (black boxes; Fig. 7A), demonstrated by bootstrap-supported changes in the nearest neighbour in phylogenetic trees. Two putative recombinant sequences (JQ837724 : HRV-A1 and JN621245 : HRV-A67) represent contemporary isolates, which form a bootstrap-supported grouping distant from other members of the same type. Within HRV-B 5'UTR sequences (Fig. 7B), only three putative recombinant types were noted, and two consisted of a single sequence, which was the sole representative of the type in question.

Discussion

Identification of novel HRV strains

Novel strains of HRV representing new genotypically assigned types have been described frequently during genetic analyses of HRV variants identified during routine

Fig. 7 **A**: Neighbour-joining phylogenetic trees constructed using HRV-A 5'UTR sequences that were >90 % complete from 167-626 (numbered according to FJ445111), compared to the corresponding VP1 region. Putative recombinant sequences are marked by a black line. **B**: Neighbour-joining phylogenetic trees constructed using HRV-B 5'UTR sequences that were >90 % complete from 167-626 (numbered according to FJ445111), compared to the corresponding VP1 region. Putative recombinant sequences are marked by a black line

clinical screening [6, 32, 52, 67]. This suggests that there is still potential for as yet undiscovered genetic variation within all three species of HRV. Consistent with these previous studies, our analysis identified 13 putative new types of HRV-B and HRV-C. These putative new types have been submitted for consideration to the Picornavirus Study Group and have been assigned new type numbers.

Estimation of mean substitution rates and analysis of pairwise p-distances

This study represents the first comparison of mean substitution rates between distant genome regions of HRV-A and HRV-C. The estimates of mean substitution rates obtained in both regions correlate with other published estimates for single-strand RNA viruses [9, 18]. Most previously documented substitution rate analysis for picornaviruses has focused on the capsid coding region, particularly VP1 [5, 16]. In one comparison of several picornaviruses, it was found that members of the genus *Enterovirus* (EV) had significantly higher mean substitution rates than other members of the family *Picornaviridae* (including members of the genera *Aphthovirus, Teschovirus*, *Hepatovirus* and *Cardiovirus*), with very limited overlap of the 95 % HPD intervals between the two groups [16]. Our estimates for both HRV-A and HRV-C showed overlapping 95 % HPD intervals with EV and non-EV examples cited in this paper. Two studies of the evolution of serotypes of EV have included estimates of substitution rates in the 3Dpol region [16, 41], and these are of the same order of magnitude as our estimates for HRV-A and HRV-C.

Similar to our observations for HRV-A and HRV-C, a general congruency of substitution rates between VP1 and 3Dpol has previously been observed in the evolution of recombinant groups of EV71 [40] and recombinant groups of three HEV-B serotypes: E9, E11 and E30 [41].

As has previously been observed in analysis of EV71 isolates [40], similar mean rates of nucleotide substitution between two distant genome regions should lead to a distribution of pairwise p-distances in these two regions with a gradient of 1 and a y-intercept of 0. This suggests that any discontinuities or outlying comparisons in distributions of pairwise p-distances in these two regions relate to potential recombination events (coloured red and yellow in Fig. 4). Outlying data points were typically the result of comparison with variants that displayed bootstrap-supported changes in tree topology in the two regions. The most striking example of this is a large cluster of recombinant sequences between 0.23 and 0.34 pairwise nucleotide p-distance in VP1 for HRV-A (coloured red in Fig. 4), which correspond to the HRV-A type groups in clade 1 (Fig. 1) that undergo a bootstrap-supported change in tree topology between VP1 and 3Dpol.

Analysis of recombination in HRV-A, HRV-B and HRV-C

A small number of recombination events have been hypothesized to have played a role in the formation of a number of HRV-A and HRV-B types [49]. More speculatively, it has also been suggested that the entire HRV-B species was formed by recombination of HEV with HRV-A [64].

In common with one previous analysis [49], we found HRV-A to have the highest number of recombinant sequences among circulating strains, with only very infrequently detected recombination events within the coding regions of HRV-B and HRV-C.

HRV-A sequences are divided into four bootstrap-supported clades, which were consistently observed in the regions analysed in this study (VP4/VP2, VP1 and 3Dpol). Further analysis indicated that these distinct clades were, in fact, present in every gene region, except for the relatively short and highly conserved VP4 and 3B regions. Recombination appears to have occurred relatively frequently within clades 1-3 (Fig. 1). However, there was no evidence of recombination occurring between the clades within any coding region, and therefore, we would suggest that these HRV-A clades are potentially now diverging on distinct and evolutionarily separate paths.

Within these four clades, most evidence of recombination involves full HRV type groups, as opposed to individual sequences. This would suggest that these events are likely to have taken place during the process of diversification into genetically distinct types.

One previous study, considering only one example from most HRV types, detected 23 recombinant genomes resulting from 12 different recombination events [49]. Our analysis identified events that are analogous to most of these. However, as we used bootstrap-supported phylogenetic trees to define nearest neighbours in different coding regions, our estimates of recombination parents occasionally differed (Table 2). As the detailed phylogeny of nearest-neighbour relationships are not well resolved, particularly in the non-structural protein-coding regions, we felt it was useful to consider the nearest bootstrap supported clade members as representing nearest-neighbour groups. Additionally, in some cases, a recombinant sequence behaves as an outgroup to a new clade, and therefore, we have treated the entire new clade as the single nearest neighbour. This approach was facilitated by the use of GroupScan to confirm recombinant sequences and to determine putative recombination breakpoints. In preference to bootscanning, this method determines the extent of grouping of a query sequence within entire pre-defined groups, rather than losing potentially phylogenetically informative sequence data by reducing control groups to

single consensus sequences. This eliminates the issue of bias of results when a query sequence is genetically divergent from both control groups [60].

Our results also affirm the previously reported finding that contemporary HRV-A76 sequences have undergone a recombination event with HRV-A56 [65]. In addition to the originally reported single recombinant sequence, we have noted that all contemporary isolates group together and therefore that the recombinant group of contemporary HRV-A76 sequences is likely to be the present dominant lineage. This pattern was observed in several other HRV types, indicating that while contemporary recombination occurring within the time period of our dataset (around 50 years) is not frequent, it can occur.

In contrast to the extensive phylogenetic incongruity observed in HRV-A, HRV-B and HRV-C show relatively infrequent or entirely absent recombination within the coding region. One putative recombinant sequence, HQ123444, belonging to the HRV-B species has been reported previously [32]. However, the previous analysis used different methods for determining recombinant sequences and did not detect the 100 % identity of this sequence with HRV-B35 sequences in the 3' end of 3Dpol, instead noting an association with HRV-B35 throughout. In fact, this sequence is identical to published strains from position 6353 onwards. One might speculate on possible explanations for this observation, as no contemporary HRV-B35 full genomes are available for comparison. It is possible that this putative recombinant may also represent a sequencing or assembly error.

In accordance with our previous study, there is a striking lack of recombination within HRV-C coding regions [38]. However, in both cases, HRV-C analysis is limited by the lack of strains older than 15 years. In HRV-A and HRV-B, all instances of contemporary sequences grouping separately from older strains involve a sequenced example that was originally collected over 40 years ago. It is therefore possible that HRV-C sequences undergo recombination at similar frequencies, but the current available time span of sequences renders their detection problematic.

Recombination within the 5'UTR of HRVs

Within HRV-A sequences, the 5'UTR was the only region analysed in which clades 1 to 4 were not maintained. We noted considerably more recombination events within the 5'UTR of HRV-A than were seen in previous studies [49]. This included two instances of contemporary sequences grouping distantly from other members of their type group. The discrepancy in results could be due to different alignment methods and the fact that we relied upon comparisons of robust groupings within phylogenetic trees to determine recombinant sequences. In addition, several putative recombinant sequences that were identified previously displayed inadequate bootstrap support on phylogenetic trees (Fig. 7A).

Recombination involving the 5'UTR of HRV-B was relatively rare, and several previously reported putative recombinants showed no bootstrap support [49]. However, the 5'UTR region of HRV-B has not been extensively sequenced to date, and only nine HRV-B types were represented by more than one sequence.

Similarly to EV [55], HRV showed evidence of inter-species recombination within the 5'UTR region. The majority of analysed HRV-C 5'UTR sequences had a bootstrap-supported grouping within the HRV-A clade (Fig. S3: Supplementary Data). HRV-C sequences were not interspersed evenly throughout the HRV-A clade, but all fell within three distinct clades. These findings are in accordance with other reports of inter-species recombination in this region [17, 38].

Analysis of recombination breakpoints within the coding region

Recombination breakpoints were determined for all putative recombinant HRV-A sequences with well-supported nearest-neighbour groups (Table 2). Similarly to EV [30, 34] and HPeV [2], the majority of breakpoints are near the P1/P2 junction. Coupled with the observation of high within-species divergence within the capsid coding region, this suggests that there may be a biological compatibility barrier that decreases the reproductive fitness of progeny that are recombinant within the capsid region. Therefore, although inter-typic recombination may occur throughout the full length of the genome in the course of a natural co-infection, recombination within the capsid region may not be represented in clinical sampling.

This study demonstrates that, despite a close phylogenetic relationship to EV, recombination within all three species of HRV was substantially less frequent than that observed within EV species. In addition, there was only very limited evidence of recombination occurring within the 40-50 years that the dataset spanned. Following upon previous reports that have provided some glimpses of the occurrence of recombination in HRV-A [49, 65], our study represents a comprehensive analysis of recombination in all published HRV full genome sequences to date, an additional analysis of HRV sequence fragments obtained from isolates spanning a 14-year time period and, to our knowledge, the first attempt to catalogue all currently detectable recombination events and potential breakpoints within HRV-A.

Acknowledgments The authors are grateful to Eleanor Gaunt for valuable assistance received in the preliminary stages of this work.

The authors are also grateful to Heli Harvala, Kate Templeton, Peter McCullough, Julie White, Mary Notman, Eleanor Leslie and Carol Thomson for providing samples, data and other virus testing results from the respiratory sample archive. Chloe McIntyre's PhD studentship was funded by the Medical Research Council.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Arden KE, McErlean P, Nissen MD, Sloots TP, Mackay IM (2006) Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections. J Med Virol 78:1232–1240
2. Benschop KS, de Vries M, Minnaar RP, Stanway G, van der Hoek L, Wolthers KC, Simmonds P (2010) Comprehensive full-length sequence analyses of human parechoviruses: diversity and recombination. J Gen Virol 91:145–154
3. Blomqvist S, Savolainen C, Råman L, Ra L, Roivainen M, Hovi T (2002) Human Rhinovirus 87 and Enterovirus 68 represent a unique serotype with rhinovirus and enterovirus features. J Clin Micro 40:4218–4223
4. Bochkov YA, Palmenberg AC, Lee W-M, Rathe JA, Amineva SP, Sun X, Pasic TR, Jarjour NN, Liggett SB, Gern JE (2011) Molecular modeling, organ culture and reverse genetics for a newly identified human rhinovirus C. Nat Med 17:627–632
5. Calvert J, Chieochansin T, Benschop KS, McWilliam Leitch EC, Drexler JF, Grywna K, da Costa Ribeiro H, Drosten C, Harvala H, Poovorawan Y, Wolthers KC, Simmonds P (2010) Recombination dynamics of human parechoviruses: investigation of type-specific differences in frequency and epidemiological correlates. J Gen Virol 91:1229–1238
6. Chidlow GR, Laing IA, Harnett GB, Greenhill AR, Phuanukoonnon S, Siba PM, Pomat WS, Shellam GR, Smith DW, Lehmann D (2012) Respiratory viral pathogens associated with lower respiratory tract disease among young children in the highlands of Papua New Guinea. J Clin Virol 54:235–239
7. Cooney MK, Fox JP, Kenny GE (1982) Antigenic groupings of 90 rhinovirus serotypes. Infect Immun 37:642–647
8. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214
9. Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9:267–276
10. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573–582
11. Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol 52:696–704
12. Hamparian VV, Colonno RJ, Cooney MK, Dick EC, Gwaltney JM, Hughes JH, Jordan WS, Kapikian AZ, Mogabgab WJ, Monto A (1987) A collaborative report: rhinoviruses–extension of the numbering system from 89 to 100. Virology 159:191–192
13. Han TH, Chung JY, Hwang ES, Koo JW (2009) Detection of human rhinovirus C in children with acute lower respiratory tract infections in South Korea. Arch Virol 154:987–991
14. Harvala H, McIntyre CL, McLeish NJ, Kondracka J, Palmer J, Molyneaux P, Gunson R, Bennett S, Templeton K, Simmonds P (2012) High detection frequency and viral loads of human rhinovirus species A to C in fecal samples; diagnostic and clinical implications. J Med Virol 84:536–542
15. Harvala H, Robertson I, McWilliam Leitch EC, Benschop K, Wolthers KC, Templeton K, Simmonds P (2008) Epidemiology and clinical associations of human parechovirus respiratory infections. J Clin Microbiol 46:3446–3453
16. Hicks AL, Duffy S (2011) Genus-Specific Substitution Rate Variability among Picornaviruses. J Virol 85:7942–7947
17. Huang T, Wang W, Bessaud M, Ren P, Sheng J, Yan H, Zhang J, Lin X, Wang Y, Delpeyroux F, Deubel V (2009) Evidence of recombination and genetic diversity in human rhinoviruses in children with acute respiratory infection. PLoS One 4:e6355
18. Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol 54:156–165
19. Kapikian A, Conant R, Hamparian V (1967) Rhinoviruses: a numbering system. Nature 213:761–762
20. Khetsuriani N, Lu X, Teague WG, Kazerouni N, Anderson LJ, Erdman DD (2008) Novel human rhinoviruses and exacerbation of asthma in children. Emerg Infect Dis 14:1793–1796
21. Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, Schnurr D, Ganem D, DeRisi JL, Boushey HA (2007) Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. J Infect Dis 196:817–825
22. Kistler AL, Webster DR, Rouskin S, Magrini V, Credle JJ, Schnurr DP, Boushey HA, Mardis ER, Li H, DeRisi JL (2007) Genome-wide diversity and selective pressure in the human rhinovirus. Virol J 4:40
23. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901
24. Kusel MM, de Klerk NH, Holt PG, Kebadze T, Johnston SL, Sly PD (2006) Role of respiratory viruses in acute upper and lower respiratory tract illness in the first year of life: a birth cohort study. Pediatr Infect Dis J 25:680–686
25. Laine P, Savolainen C, Blomqvist S, Hovi T (2005) Phylogenetic analysis of human rhinovirus capsid protein VP1 and 2A protease coding sequences confirms shared genus-like relationships with human enteroviruses. J Gen Virol 86:697–706
26. Lamson D, Renwick N, Kapoor V, Liu Z, Palacios G, Ju J, Dean A, St George K, Briese T, Lipkin WI (2006) MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004–2005. J Infect Dis 194:1398–1402
27. Ledford RM, Patel NR, Demenczuk TM, Watanyar A, Herbertz T, Collett MS, Pevear DC, Irol JV (2004) VP1 Sequencing of All Human Rhinovirus Serotypes : Insights into Genus Phylogeny and Susceptibility to Antiviral Capsid-Binding Compounds. J Virol 78:3663–3674
28. Lee WM, Kiesner C, Pappas T, Lee I, Grindle K, Jartti T, Jakiela B, Lemanske RF, Shult PA, Gern JE (2007) A diverse group of previously unrecognized human rhinoviruses are common causes of respiratory illnesses in infants. PLoS One 3:e966
29. Lewis-Rogers N, Bendall ML, Crandall KA (2009) Phylogenetic relationships and molecular adaptation dynamics of human rhinoviruses. Mol Biol Evol 26:969–981
30. Lindberg AM, Andersson P, Savolainen C, Mulders MN, Hovi T (2003) Evolution of the genome of Human enterovirus B: incongruence between phylogenies of the VP1 and 3CD regions indicates frequent recombination within the species. J Gen Virol 84:1223–1235
31. Linsuwanon P, Payungporn S, Samransamruajkit R, Posuwan N, Makkoch J, Theanboonlers A, Poovorawan Y (2009) High prevalence of human rhinovirus C infection in Thai children with acute lower respiratory tract disease. J Infect 59:115–121

32. Linsuwanon P, Payungporn S, Suwannakarn K, Chieochansin T, Theamboonlers A, Poovorawan Y (2011) Complete coding sequence characterization and comparative analysis of the putative novel human rhinovirus (HRV) species C and B. Virol J 8:5
33. Lukashev AN, Lashkevich VA, Ivanova OE, Koroleva GA, Hinkkanen AE, Ilonen J (2003) Recombination in Circulating Enteroviruses. J Virol 77:10423–10431
34. Lukashev AN, Lashkevich VA, Ivanova OE, Koroleva GA, Hinkkanen AE, Ilonen J (2005) Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. J Gen Virol 86:3281–3290
35. Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21:98–102
36. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462–2463
37. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563
38. McIntyre CL, McWilliam Leitch EC, Savolainen-Kopra C, Hovi T, Simmonds P (2010) Analysis of genetic diversity and sites of recombination in human rhinovirus species C. J Virol 84:10297–10310
39. McWilliam Leitch EC, Bendig J, Cabrerizo M, Cardosa J, Hyypiä T, Ivanova OE, Kelly A, Kroes AC, Lukashev A, MacAdam A, McMinn P, Roivainen M, Trallero G, Evans DJ, Simmonds P (2009) Transmission networks and population turnover of echovirus 30. J Virol 83:2109–2118
40. McWilliam Leitch EC, Cabrerizo M, Cardosa J, Harvala H, Ivanova OE, Koike S, Kroes AC, Lukashev AN, Perera D, Roivainen M, Susi P, Trallero G, Evans DJ, Simmonds P (2012) The association of recombination events in the founding and emergence of subgenogroup evolutionary lineages of human enterovirus 71. J Virol 86:2676–2685
41. McWilliam Leitch EC, Cabrerizo M, Cardosa J, Harvala H, Ivanova OE, Kroes AC, Lukashev A, Muir P, Odoom J, Roivainen M, Susi P, Trallero G, Evans DJ, Simmonds P (2010) Evolutionary dynamics and temporal/geographical correlates of recombination in the human enterovirus echovirus types 9, 11, and 30. J Virol 84:9292–9300
42. McWilliam Leitch EC, Harvala H, Robertson I, Ubillos I, Templeton K, Simmonds P (2009) Direct identification of human enterovirus serotypes in cerebrospinal fluid by amplification and sequencing of the VP1 region. J Clin Microbiol 44:119–124
43. Miller EK, Edwards KM, Weinberg GA, Iwane MK, Griffin MR, Hall CB, Zhu Y, Szilagyi PG, Morin LL, Heil LH, Lu X, Williams JV (2009) A novel group of rhinoviruses is associated with asthma hospitalizations. J Allergy Clin Immunol 123(98–104):e1
44. Oberste MS, Maher K, Pallansch MA (2004) Evidence for frequent recombination within species human enterovirus B based on complete genomic sequences of all thirty-seven serotypes. J Virol 78:855–867
45. Oberste MS, Peñaranda S, Pallansch MA (2004) RNA recombination plays a major role in genomic change during circulation of coxsackie B viruses. J Virol 78:2948–2955
46. Oberste MS, Maher K, Kilpatrick DR, Flemister MR, Brown BA, Pallansch MA (1999) Typing of human enteroviruses by partial sequencing of VP1. J Clin Microbiol 37:1288–1293
47. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225
48. Palmenberg AC, Rathe JA, Liggett SB (2010) Analysis of the complete genome sequences of human rhinovirus. J Allergy Clin Immunol 125:1190–1199
49. Palmenberg AC, Spiro D, Kuzmickas R, Wang S, Djikeng A, Rathe JA, Fraser-Liggett CM, Liggett SB (2009) Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. Science 324:55–59
50. Peltola V, Waris M, Osterback R, Susi P, Ruuskanen O, Hyypiä T (2008) Rhinovirus transmission within families with children: incidence of symptomatic and asymptomatic infections. J Infect Dis 197:382–389
51. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci USA 98:13757–13762
52. Rathe JA, Liu X, Tallon LJ, Gern JE, Liggett SB (2010) Full-genome sequence and analysis of a novel human rhinovirus strain within a divergent HRV-A clade. Arch Virol 155:83–87
53. Renwick N, Schweiger B, Kapoor V, Liu Z, Villari J, Bullmann R, Miething R, Briese T, Lipkin WI (2007) A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. J Infect Dis 196:1754–1760
54. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
55. Santti J, Hyypiä T, Kinnunen L, Salminen M (1999) Evidence of recombination among enteroviruses. J Virol 73:8741–8749
56. Savolainen C, Blomqvist S, Mulders MN, Hovi T (2002) Genetic clustering of all 102 human rhinovirus prototype strains: serotype 87 is close to human enterovirus 70. J Gen Virol 83:333–340
57. Savolainen C, Mulders MN, Hovi T (2002) Phylogenetic analysis of rhinovirus isolates collected during successive epidemic seasons. Virus Res 85:41–46
58. Simmonds P (2012) SSE: a nucleotide and amino acid sequence analysis platform. BMC Res Notes 5:50
59. Simmonds P, McIntyre C, Savolainen-Kopra C, Tapparel C, Mackay IM, Hovi T (2010) Proposals for the classification of human rhinovirus species C into genotypically assigned types. J Gen Virol 91:2409–2419
60. Simmonds P, Midgley S (2005) Recombination in the genesis and evolution of hepatitis B virus genotypes. J Virol 79:15467–15476
61. Simmonds P, Welch J (2006) Frequency and dynamics of recombination within different species of human enteroviruses. J Virol 80:483–493
62. Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129
63. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
64. Tapparel C, Junier T, Gerlach D, Cordey S, Van Belle S, Perrin L, Zdobnov EM, Kaiser L (2007) New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features. BMC Genomics 8:224
65. Tapparel C, Junier T, Gerlach D, Van-Belle S, Turin L, Cordey S, Mühlemann K, Regamey N, Aubert J-D, Soccal PM, Eigenmann P, Zdobnov E, Kaiser L (2009) New respiratory enterovirus and recombinant rhinoviruses among circulating picornaviruses. Emerg Infect Dis 15:719–726
66. Tapparel C, L'Huillier AG, Rougemont AL, Beghetti M, Barazzone-Argiroffo C, Kaiser L (2009) Pneumonia and pericarditis in a child with HRV-C infection: a case report. J Clin Virol 45:157–160
67. Wisdom A, McWilliam Leitch EC, Gaunt E, Harvala H, Simmonds P (2009) Screening respiratory samples for detection of human rhinoviruses (HRVs) and enteroviruses: comprehensive VP4-VP2 typing reveals high incidence and genetic diversity of HRV species C. J Clin Microbiol 47:3958–3967
68. Wisdom A, Kutkowska AE, McWilliam Leitch EC, Gaunt E, Templeton K, Harvala H, Simmonds P (2009) Genetics, recombination and clinical features of human rhinovirus species C

(HRV-C) infections; interactions of HRV-C with other respiratory viruses. PLoS One 4:e8518
69. Xatzipsalti M, Kyrana S, Tsolia M, Psarras S, Bossios A, Laza-Stanca V, Johnston SL, Papadopoulos NG (2005) Rhinovirus viremia in children with respiratory infections. Am J Respir Crit Care Med 172:1037–1040
70. de Almeida MB, Zerbinati RM, Tateno AF, Oliveira CM, Romão RM, Rodrigues JC, Pannuti CS, da Silva Filho LV (2010) Rhinovirus C and respiratory exacerbations in children with cystic fibrosis. Emerg Infect Dis 16:996–999

Journal of General Virology (2010), **91**, 2409–2419 DOI 10.1099/vir.0.023994-0

Review

Proposals for the classification of human rhinovirus species C into genotypically assigned types

Peter Simmonds,[1] Chloe McIntyre,[1] Carita Savolainen-Kopra,[2] Caroline Tapparel,[3] Ian M. Mackay[4] and Tapani Hovi[2]

Correspondence
Peter Simmonds
peter.simmonds@ed.ac.uk

[1]Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, UK

[2]National Institute for Health and Welfare, Mannerheimintie 166, FIN-00300 Helsinki, Finland

[3]Central Laboratory of Virology, University of Geneva Hospitals, 1211 Geneva 14, Switzerland

[4]Emerging Viruses Group, Queensland Paediatric Infectious Diseases Laboratory, Department of Infectious Diseases, Royal Children's Hospital, Queensland Children's Medical Research Institute, The University of Queensland, Brisbane, Queensland 4029, Australia

Human rhinoviruses (HRVs) are common respiratory pathogens associated with mild upper respiratory tract infections, but also increasingly recognized in the aetiology of severe lower respiratory tract disease. Wider use of molecular diagnostics has led to a recent reappraisal of HRV genetic diversity, including the discovery of HRV species C (HRV-C), which is refractory to traditional virus isolation procedures. Although it is heterogeneous genetically, there has to date been no attempt to classify HRV-C into types analogous to the multiple serotypes identified for HRV-A and -B and among human enteroviruses. Direct investigation of cross-neutralization properties of HRV-C is precluded by the lack of methods for *in vitro* culture, but sequences from the capsid genes (VP1 and partial VP4/VP2) show evidence for marked phylogenetic clustering, suggesting the possibility of a genetically based system comparable to that used for the assignment of new enterovirus types. We propose a threshold of 13 % divergence for VP1 nucleotide sequences for type assignment, a level that classifies the current dataset of 86 HRV-C VP1 sequences into a total of 33 types. We recognize, however, that most HRV-C sequence data have been collected in the VP4/VP2 region (currently 701 sequences between positions 615 and 1043). We propose a subsidiary classification of variants showing >10 % divergence in VP4/VP2, but lacking VP1 sequences, to 28 provisionally assigned types (subject to confirmation once VP1 sequences are determined). These proposals will assist in future epidemiological and clinical studies of HRV-C conducted by different groups worldwide, and provide the foundation for future exploration of type-associated differences in clinical presentations and biological properties.

Introduction

Human rhinoviruses (HRVs) are highly prevalent respiratory pathogens, most commonly associated with mild upper respiratory tract disease and exacerbations of pre-existing respiratory disease such as asthma. They are also increasingly recognized as underlying more severe disease manifestations, such as bronchiolitis in young children and in the immunosuppressed. The increasing use of molecular methods for respiratory virus screening has contributed to this reappraisal of rhinoviruses, as has the recent discovery of an entirely novel rhinovirus group, refractory to previously used virus isolation methods but now known to be highly prevalent and widely circulating worldwide (Arden *et al.*, 2006; Kaiser *et al.*, 2006; Lamson *et al.*, 2006; Kistler *et al.*, 2007; Lau *et al.*, 2007; Lee *et al.*, 2007; McErlean *et al.*, 2007; Renwick *et al.*, 2007; Olenec *et al.*, 2010).

These newly characterized rhinoviruses have been proposed to belong to a novel species of rhinovirus (designated species C; HRV-C), recognizing their substantial sequence divergence from other classified species within the genus *Enterovirus* of picornaviruses (Carstens, 2010; Knowles, 2010) (Fig. 1a). Clinically and biologically, they share many attributes with the other designated HRV species, HRV-A and -B. Most studies of HRV-C disease associations, typically focused on children from asthmatic and/or hospital-based populations (Arden & Mackay, 2010), have demonstrated a similarly broad range of clinical outcomes to those observed in HRV-A and -B infections and, indeed, with other respiratory viruses. Some

A full list of all HRV-C variants characterized to date, categorized into a total of 61 confirmed or provisionally assigned types, is available with the online version of this paper.

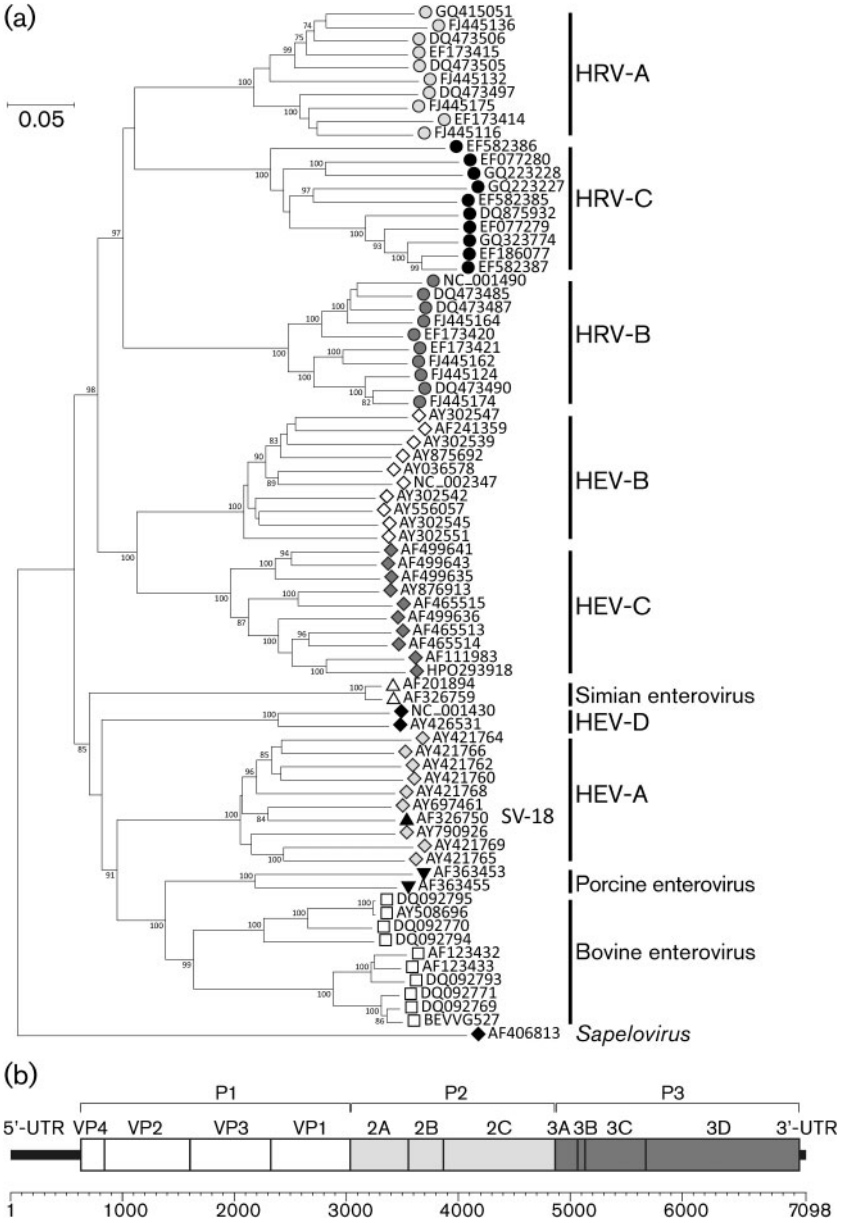


Fig. 1. (a) Sequence relationships between species currently assigned to the genus *Enterovirus*, depicted by phylogenetic analysis of aligned sequences from the P1 (capsid protein-encoding) region [positions 616–3125, numbered according to the reference sequence (GenBank accession no. EF582385); species identified by different symbols and shading]. The tree was constructed by neighbour-joining analysis of pairwise amino acid *p* distances, with branches showing $\geqslant$70 % bootstrap support labelled. The sequence of porcine enterovirus 8 (genus *Sapelovirus*) was used as an outgroup. For presentation purposes, only the 10 most divergent sequences were included for species containing more than 10 sequences (HEV-A, -B, -C, HRV-A, -B and bovine enteroviruses). Bar, amino acid *p* distance of 0.05. (b) Diagram of the rhinovirus genome, identifying the 5′- and 3′-UTRs and structural (P1) and non-structural (P2, P3) gene regions, along with the designations of their encoded proteins. The genome is drawn to scale, using the complete genome sequence 024 (GenBank accession no. EF582385) for numbering.

studies show no difference in clinical outcome between HRV species (Lau *et al.*, 2007; Piotrowska *et al.*, 2009), whereas others provide evidence for a more frequent role of HRV-C in lower respiratory tract disease, febrile wheeze in infants and toddlers, and asthma exacerbations in older children (Lau *et al.*, 2007; Khetsuriani *et al.*, 2008; Miller *et al.*, 2009a; Wisdom *et al.*, 2009b). Contrastingly, one study described a shorter duration of asthma symptoms and less cough than seen in HRV-A infection (Arden *et al.*, 2010a).

HRV-C shares a number of features of its genomic organization with other members of the genus *Enterovirus* (Fig. 1b). This includes an approximately 7100 base genome containing a single reading frame, the absence of a leader protein, a P1 region encoding four capsid proteins, a 2A gene encoding a *cis*-acting proteinase, followed by a series of non-structural proteins collinear with those of other enteroviruses, including 3D (which encodes the RNA-dependent RNA polymerase). HRV-C contains a type I internal ribosomal entry site that is structurally similar to and has short regions of striking sequence conservation with those of other enteroviruses. Members of the genus *Enterovirus* do, however, differ in other aspects of their genome organization. Most evident is the variability in the position of the *cis*-acting replication element. This is located at a homologous position within the 2C coding sequence in all four human enterovirus (HEV) species A–D, but is variable in position in each rhinovirus species [within 2A in HRV-A (Gerber *et al.*, 2001), VP1 in HRV-B and proposed to be located at the 5′ end of VP2 in species C (Cordey *et al.*, 2008)]. In marked contrast to the ease with which HRV-A and -B can be isolated, HRV-C has, to date based on the

published literature, been unculturable *in vitro* (Lau *et al.*, 2007; McErlean *et al.*, 2007). Historically, this hindered its discovery and has additionally precluded investigation of its antigenic variability, a feature characteristic of other HRVs and HEVs.

Proposal aims

Genetic characterization of HRV-C amplified from clinical specimens has provided evidence for extensive heterogeneity in the VP4/VP2 region, the existence of two phylogenetically separate groups of sequences in the 5′-UTR (one resembling sequences found in species A rhinoviruses; Han *et al.*, 2009; Huang *et al.*, 2009; Savolainen-Kopra *et al.*, 2009b; Wisdom *et al.*, 2009a) and substantial sequence divergence throughout the genome of the 11 full-length HRV-C sequences obtained to date (Lamson *et al.*, 2006; Lau *et al.*, 2007; Kistler *et al.*, 2007; McErlean *et al.*, 2007; Huang *et al.*, 2009; Tapparel *et al.*, 2009b; Arden *et al.*, 2010b). Although we currently lack the means to classify HRV-C serologically (as has been achieved for other rhinovirus and enterovirus species), we recognize and are responding to the need to develop a classification system for this species. This will assist in organizing the rapidly accumulating sequence data currently being generated from virological and clinical studies, and allow assignment of uniform type descriptions that will enable comparison of genetic variants characterized in separate studies over time and across different geographical regions.

This process is potentially made easier by the evident similarities in the pattern of sequence divergence of HRV-C to other rhinoviruses and enteroviruses for which classification methods have been developed and standardized. For example, the large number of distinct genetic lineages identifiable by sequence comparisons in the VP4/VP2 region matches the diversity in this region shown by different serotypes of species A and B rhinoviruses (Lau *et al.*, 2009). These differences are mirrored in VP1 (Huang *et al.*, 2009; Wisdom *et al.*, 2009b; McIntyre *et al.*, 2010) and other exposed regions of the capsid that underlie the latter's putative antigenic diversity. In developing classification proposals for HRV-C, we have strived to develop criteria for type assignment that are consistent with the principles used for other rhinoviruses and enteroviruses, whilst acknowledging differences in its diversity, genetic history and biology.

Classification of enteroviruses and rhinoviruses into serotypes

HRV-A and -B comprise a number of antigenically distinct viruses designated on the basis of their cross-neutralization properties *in vitro*, currently totalling 74 serotypes of HRV-A and 25 of HRV-B (Kapikian *et al.*, 1967). These totals incorporate minor adjustments to the original classification when HRV was subsequently characterized genetically (Kapikian *et al.*, 1971; Hamparian *et al.*, 1987). As examples, one of the classical HRV prototypes, HRV-87, was found to belong to the species HEV-D as a variant of HEV-68 (Blomqvist *et al.*, 2002). A strain referred to as Hanks was considered to represent a candidate new type

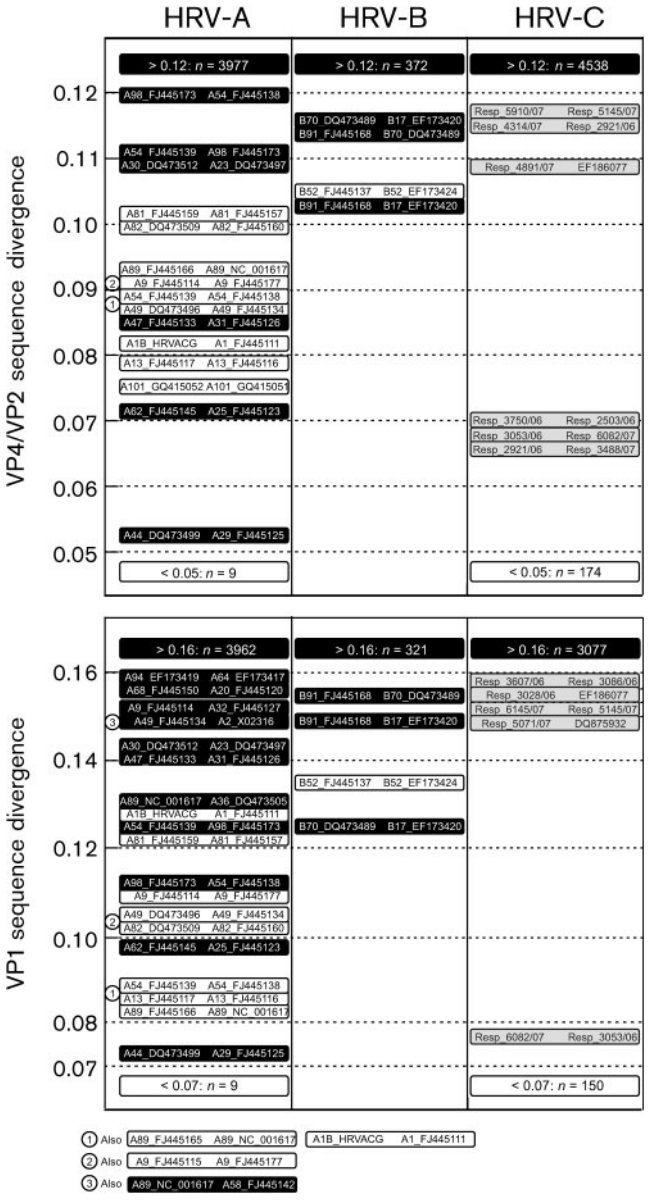


Fig. 2. Distributions of pairwise nucleotide *p* (uncorrected) distances (*y*-axis) between HRV-A and -B variants around the previously proposed thresholds (Savolainen *et al.*, 2002) dividing inter- and intra-serotype distances in VP1 (lower panel) and VP4/VP2 (upper panel). Pairwise comparisons in the equivalent distance range for HRV-C are shown for comparison. Unfilled boxes indicate pairwise distances between HRV variants previously classified as the same serotype; black-filled boxes, pairwise comparisons between different serotypes; grey boxes, pairwise comparisons of variants with unknown serological cross-reactivity (HRV-C). For clarity, multiple examples of pairwise distances between the same (sero)type pairs have not been shown.

but, on sequence analysis, it was found to be genetically similar to HRV-21 (Ledford *et al.*, 2004; Laine *et al.*, 2005).

In common with HEVs (Oberste *et al.*, 1999), there is a close correlation between sequence divergence of HRV-A and -B in the VP1 region (and other structural genes) and their designated serotypes (Savolainen *et al.*, 2002; Ledford *et al.*, 2004; Laine *et al.*, 2005). For HEVs, a nucleotide sequence divergence value of >25 % in VP1 (>15 % amino acid sequence difference) may be used as an alternative means to classify more recently discovered types without recourse to extensive serological characterization (Oberste *et al.*, 1999). Application of this principle has led to the assignment of 40 genotypically defined enterovirus 'types' in addition to the 64 traditionally classified serotypes. Thresholds of 12 % similarly differentiate inter- from intra-serotype divergences in the VP1 gene of species A and B rhinoviruses, respectively (Savolainen *et al.*, 2002; McIntyre *et al.*, 2010), providing the means in principle to detect novel species A types (e.g. Wisdom *et al.*, 2009b) without assaying for cross-neutralization (Savolainen *et al.*, 2002).

For rhinoviruses it is, however, recognized that some inconsistencies and overlap of divergence values between and within serotypes exist (Savolainen *et al.*, 2002). In species A, pairwise divergence values in VP1 between serotypes 95 and 8 (1.6 %), serotypes 44 and 29 (7.3 %), serotypes 62 and 25 (9.4 %) and serotypes 98 and 54 (11.4 %) are interspersed with those observed within serotypes, as is the pairwise distance between the species B serotypes 70 and 17 (12.3 %) (Fig. 2, lower panel). Overlaps in inter- and intra-serotype distances, often involving the same serotype pairs, are also evident from an equivalent analysis of pairwise distances of VP4/VP2 sequences (Fig. 2, upper panel). In two cases (95/8, 44/29), reanalysis demonstrated that these pairs did indeed show cross-neutralization (Cooney *et al.*, 1982; Ledford *et al.*, 2004), whereas one of the more divergent pairs, 62/25, did not (Cooney *et al.*, 1982). There is clearly scope to reinvestigate cross-reactivity of the other discrepant pairs. Overall, however, the actual distributions of pairwise distances between different serotypes of HEVs and rhinoviruses overlap minimally with intra-serotype nucleotide distances (as exemplified by HEV-B and HRV-A; Fig. 3a, b). For enteroviruses, the lowest value between these two distributions closely matches the type-assignment threshold now used for genotypic assignment of new

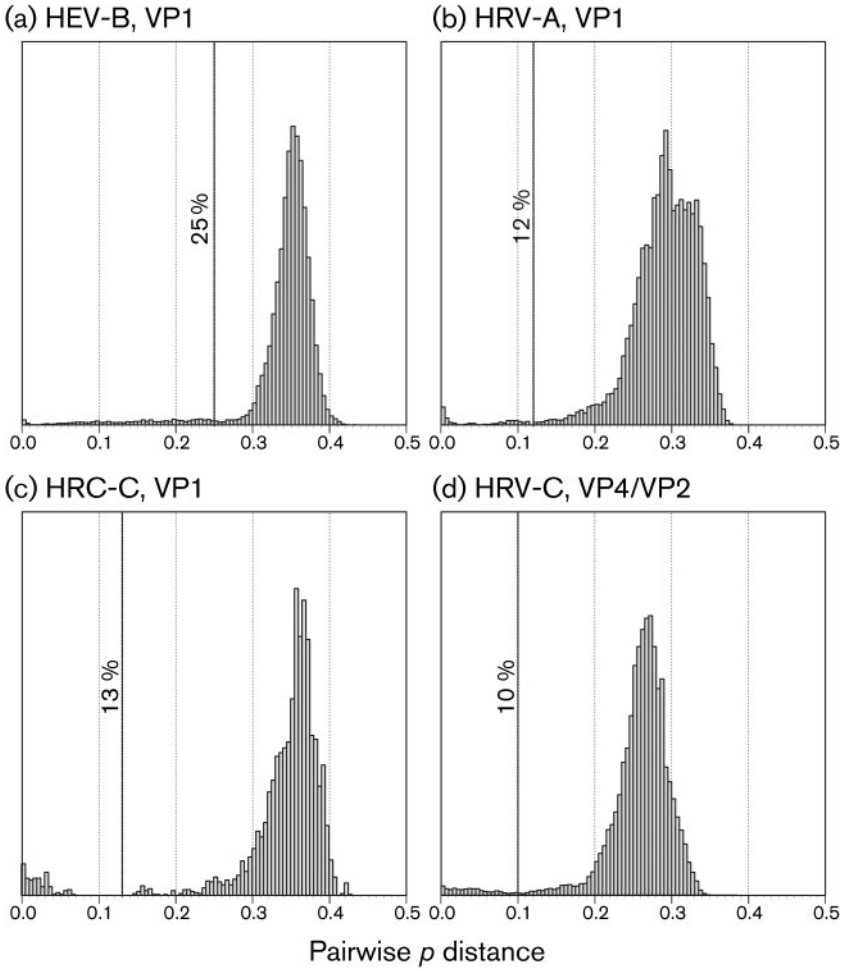


Fig. 3. Distribution of pairwise nucleotide *p* distances between: (a) available complete (>90 %) VP1 sequences of HEV-B (n=330); (b) available complete (>90 %) VP1 sequences of HRV-A (n=279); (c) available VP1 sequences of HRV-C (n=86; >90 % complete between positions 2304 and 3125); (d) all available VP4/VP2 sequences of HRV-C (n=702; >90 % complete between positions 615 and 1043). Previously designated (HEV) or proposed type thresholds are indicated by thick lines.

types. We propose to adopt this method for a genotypic classification of HRV-C.

Rhinovirus recombination

The process of recombination creates chimaeric virus genomes in which different genome regions have separate evolutionary origins; recombinants may change in their phylogenetic relationships to other sequences between genome regions. For enteroviruses and rhinoviruses whose type assignments are dependent on the capsid genes (and the encoded differences in antigenic properties), regions that undergo extremely frequent recombination (such as the 5′-UTR, P2 and P3 non-structural gene regions in HEV) cannot therefore contribute to their (sero)type classification (Savolainen-Kopra *et al.*, 2009b).

Rhinovirus species A and B show much more consistent phylogeny relationships across the genome, as exemplified by the largely concordant phylogenies of the 3Dpol and VP4/VP2 regions (Savolainen *et al.*, 2004). There are, however, some inconsistencies evident on analysis of complete genome sequences of species A and B (Palmenberg *et al.*, 2009; Tapparel *et al.*, 2009a). For example, HRV-53 shows greater similarity to HRV-46 in the non-structural region than anticipated by their sequence relationship in the capsid-encoding region, whilst a similar comparison of HRV-78 and HRV-12 showed non-structural gene sequences to be more divergent. In these and other instances, most changes in phylogenetic relationship occurred at the P1/P2 boundary, implying separate evolutionary origins for the structural and non-structural gene blocks in some serotypes.

In contrast to species A and B, our recent extensive comparison of phylogenies of the VP4/VP2, VP1 and 3Dpol regions of species C demonstrated consistent branching orders and relative branch lengths in all three coding regions (McIntyre *et al.*, 2010). However, several phylogeny violations occurred between the 5′-UTR and VP4/VP2 trees, originating from a series of likely interspecies recombination events with breakpoints towards the 3′ end of the 5′-UTR (Han *et al.*, 2009; Huang *et al.*, 2009; Savolainen-Kopra *et al.*, 2009b; Wisdom *et al.*, 2009a).

Remarkably, most 5′-UTR sequences of species C cluster within the species A 5′-UTR clade, with the remainder being phylogenetically distinct (Han *et al.*, 2009; Huang *et al.*, 2009; Savolainen-Kopra *et al.*, 2009b; Wisdom *et al.*, 2009a). Those with species A-like 5′-UTR sequences have been named HRV-Ca, with the remainder assigned as HRV-Cc (Huang *et al.*, 2009). We have recently found that the region of 2A encoding the C-terminal domain of the proteinase also has a recombinant origin, with all available HRV-C sequences from this region clustering within the HRV-A clade (McIntyre *et al.*, 2010). The evolutionary events and the selection pressures underlying these instances of HRV-A/-C interspecies recombination are currently unknown.

HRV-C heterogeneity and proposals for type assignments

Eleven (near-)complete genome sequences, 541 sequences from the VP4/VP2 region [>90 % complete between

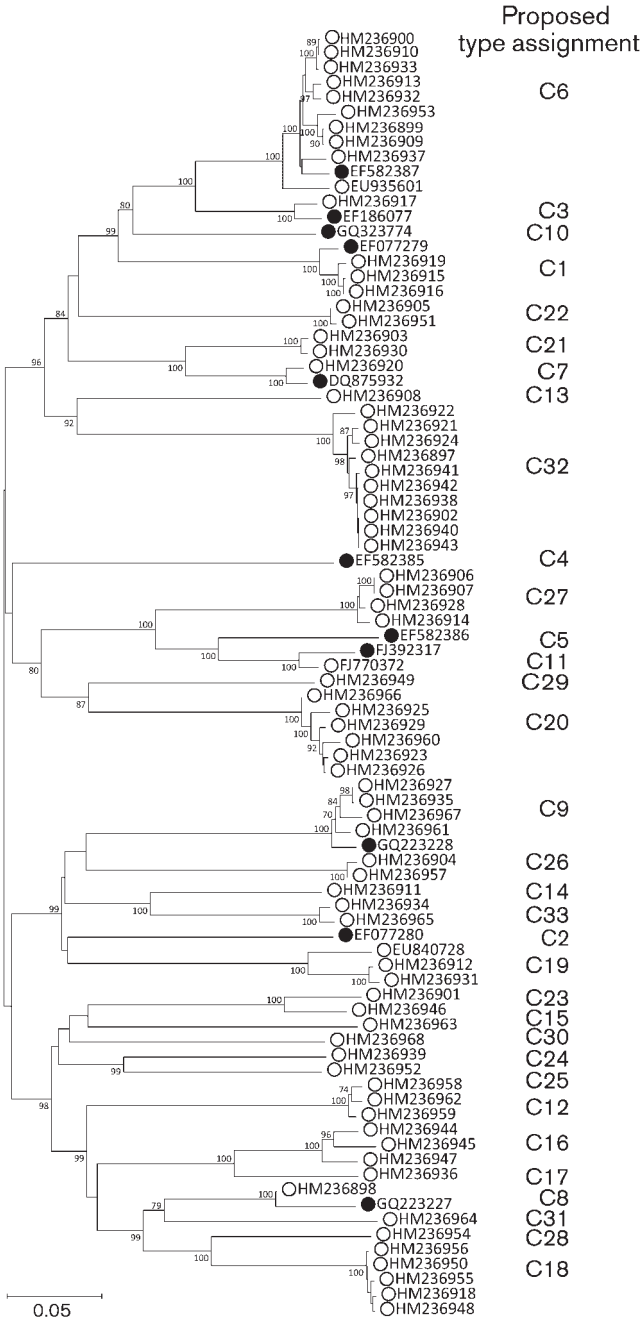


Fig. 4. Phylogenetic analysis of all available sequences of HRV-C in the VP1 region (>90 % complete between positions 2304 and 3125). The tree was constructed by neighbour-joining analysis of pairwise maximum composite likelihood distances implemented in the program MEGA (Tamura *et al.*, 2007); branches showing ⩾70 % bootstrap support are indicated. Complete genome sequences are labelled ●. Bar, maximum composite likelihood distance of 0.05.

positions 615 and 1043, numbered here and below using the complete genome HRV-C sequence 024 (GenBank accession no. EF582385)] from public databases and a further 160 unpublished sequences contributed by the authors of the current study, along with 86 complete VP1 (positions 2304–3125) and 89 partial 3Dpol (positions 6384–6854) sequences, collectively attest to the substantial genetic heterogeneity of HRV-C. The formation of a number of discrete clades of HRV-C in each genome region (Savolainen *et al.*, 2002; McIntyre *et al.*, 2010) (Fig. 4) supports the idea that genetic variants of HRV-C might be usefully classified into a number of types, comparable to types/serotypes of other HRV species.

The current lack of data on antigenic properties of HRV-C is unlikely to be addressed in the near future, due to difficulties with *in vitro* culture and the daunting task of creating and applying the necessary serotyping reagents should a viral culture system be developed. These factors necessitate a genotypic classification method. To investigate whether clear inter- and intra-type thresholds can be defined for HRV-C, we constructed a frequency histogram of the set of pairwise distances between all available sequences from the VP1 region (Fig. 3c) using previously described methods for constructing sequence alignments and determining sequence distances (McIntyre *et al.*, 2010). For comparison, we have additionally analysed an even larger dataset of available VP4/VP2 sequences (Fig. 3d).

As described previously (McIntyre *et al.*, 2010), the distribution of HRV-C VP1 sequence distances is indeed bimodal, with a clearly defined minimum (zero) value below 14.9 % and above 8 % (Fig. 2), which may be used as a threshold for putative assignment of HRV-C types. This corresponds closely to the 12 % threshold that divides within- and between-serotype distances in species A and B rhinoviruses (Savolainen *et al.*, 2002; McIntyre *et al.*, 2010). The distribution of pairwise distances in the VP4/VP2 region resembles that of VP1, with an equivalent minimum value corresponding to the type threshold of VP1 at 10 %. However, as a likely result of its shorter length and lesser degree of sequence diversity than VP1, the type threshold for VP4/VP2 was less clearly defined (Fig. 2). This pattern was also found in a similar comparison of VP4/VP2 region distance distributions in HRV-A and -B (McIntyre *et al.*, 2010) and in human enteroviruses (Oberste *et al.*, 1999; Mulders *et al.*, 2000).

Type-assignment proposals

In formulating the following criteria for type assignment, we are aware of the need for simplicity and transparency in the assignment process and the use of criteria comparable to those used for genotypic classification in other enterovirus species. At the same time, these proposals should respect and adapt to differences in the pattern of diversity in species C and the occurrence of recombination. In addition, we acknowledge that current surveillance and genetic characterization of HRV-C are incomplete and we state the need for review of and, if necessary, revision of type-assignment criteria as further genetic data become available in the future. Finally, the use of genetic comparisons in restricted regions of the genome (VP1 and VP4/VP2) should not diminish perceptions of the importance of other genomic regions in shaping the phenotype of HRV-C. However, these, together with putative biological/epidemiological differences to be found in the future, lie specifically in the realm of research enquiry and we advise against their use as subsidiary or alternative classification criteria unless or until there is a future major reappraisal of our understanding of HRV diversity and genetics.

(a) A proposed HRV-C type should be phylogenetically distinct and show >13 % nucleotide sequence divergence in VP1 from all other

Table 1. Proposed assigment of complete genome sequences of HRV-C into types

Type assignment	GenBank accession no.	Strain identifier	Submission	Variants		Reference*
				VP4/VP2	VP1	
HRV-C1	EF077279	NAT001	20 Oct 2006	17	4	Kistler *et al.* (2007)
HRV-C2	EF077280	NAT045	20 Oct 2006	36	1	Kistler *et al.* (2007)
HRV-C3	EF186077	QPM	14 Dec 2006	12	2	McErlean *et al.* (2007)
HRV-C4	EF582385	024	27 Apr 2007	3	1	Lau *et al.* (2007)
HRV-C5	EF582386	025	27 Apr 2007	22	1	Lau *et al.* (2007)
HRV-C6	EF582387	026	27 Apr 2007	32	11	Lau *et al.* (2007)
HRV-C7	DQ875932	NY-074	14 Jul 2008	4	2	Lamson *et al.* (2006)
HRV-C8	GQ223227	N4	29 May 2009	6	2	Huang *et al.* (2009)
HRV-C9	GQ223228	N10	29 May 2009	27	5	Huang *et al.* (2009)
HRV-C10	GQ323774	QCE	29 Jun 2009	7	1	Arden *et al.* (2010b)
HRV-C11	EU840952	CL-170085	21 May 2010	11	2	Tapparel *et al.* (2009b)

*Reference for first submitted sequence for each type.

Table 2. Proposed type assigment of HRV-C variants represented by VP1 (and VP4/VP2) sequences

Type assignment	VP4/VP2 region			VP1 region			References*
	GenBank accession no.	Submission date	Variants	GenBank accession no.	Submission date	Variants	
HRV-C12	EF077264	20 Oct 2006	25	HM236958	14 May 2010	3	Kistler *et al.* (2007); McIntyre *et al.* (2010)
HRV-C13	EU081795	3 Aug 2007	6	HM236908	14 May 2010	1	Renwick *et al.* (2007); McIntyre *et al.* (2010)
HRV-C14	EU081796	3 Aug 2007	6	HM236911	14 May 2010	1	Renwick *et al.* (2007); McIntyre *et al.* (2010)
HRV-C15	EU081800	3 Aug 2007	22	HM236963	14 May 2010	1	Renwick *et al.* (2007); McIntyre *et al.* (2010)
HRV-C16	EU081808	3 Aug 2007	31	HM236944	14 May 2010	3	Renwick *et al.* (2007); McIntyre *et al.* (2010)
HRV-C17	EU081809	3 Aug 2007	3	HM236936	14 May 2010	1	Renwick *et al.* (2007); McIntyre *et al.* (2010)
HRV-C18	EU590074	25 Mar 2008	30	HM236918	14 May 2010	5	Savolainen-Kopra *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C19	EU697850	5 May 2008	9	EU840728	20 Jun 2008	3	Briese *et al.* (2008); Tapparel *et al.* (2009a)
HRV-C20	EU697851	5 May 2008	14	HM236923	14 May 2010	6	Briese *et al.* (2008); McIntyre *et al.* (2010)
HRV-C21	EU752377	26 May 2008	16	HM236903	14 May 2010	2	Miller *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C22	EU752381	26 May 2008	10	HM236905	14 May 2010	2	Miller *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C23	EU752424	26 May 2008	21	HM236901	14 May 2010	2	Miller *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C24	EU752426	26 May 2008	14	HM236939	14 May 2010	1	Miller *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C25	EU752427	26 May 2008	15	HM236952	14 May 2010	1	Miller *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C26	EU752441	26 May 2008	19	HM236904	14 May 2010	2	Miller *et al.* (2009a); McIntyre *et al.* (2010)
HRV-C27	GQ223122	9 Jan 2009	14	HM236906	14 May 2010	4	Huang *et al.* (2009); McIntyre *et al.* (2010)
HRV-C28	GQ223134	9 Jan 2009	11	HM236954	14 May 2010	1	Huang *et al.* (2009); McIntyre *et al.* (2010)
HRV-C29	FJ615699	9 Jan 2009	4	HM236949	14 May 2010	1	Miller *et al.* (2009b); McIntyre *et al.* (2010)
HRV-C30	GQ476669	13 Aug 2009	2	HM236968	14 May 2010	1	Wisdom *et al.* (2009b); McIntyre *et al.* (2010)
HRV-C31	GU294380	4 Dec 2009	4	HM236964	14 May 2010	1	Wisdom *et al.* (2009b); McIntyre *et al.* (2010)
HRV-C32	GU294466	4 Dec 2009	13	HM236897	14 May 2010	10	Wisdom *et al.* (2009b); McIntyre *et al.* (2010)
HRV-C33	GU294480	4 Dec 2009	3	HM236934	14 May 2010	2	Wisdom *et al.* (2009b); McIntyre *et al.* (2010)

*References for first submitted sequence for each type.

previously classified species C types. The VP1 sequence obtained for this sequence comparison must be >90 % complete between positions 2304 and 3125 for determining valid nucleotide sequence distances. The proposed threshold corresponds to approximately 8 % amino acid sequence divergence in VP1. However, for clarity and avoidance of conflicting assignments, we do not propose amino acid distances as an additional or alternative criterion for type assignments.

(b) A sequence from the VP4/VP2 region (between positions 615 and 1043) can be used for identification of HRV-C types among the much larger dataset of VP4/VP2 sequences that have been obtained from surveillance studies.

(c) Types should be numbered sequentially from 1 using a 'C' prefix to distinguish them from serotype designations of other HRV species. In the tables of assignments drawn up, numbering commences with the 11 (near)-complete genome sequences HRV-C1 to -C11 (Table 1), based on submission date to GenBank.

(d) Subsequent assignments have been made (HRV-C12 onwards) to genetic variants of HRV-C for which VP1 and VP4/VP2 sequences are available, again ordered by submission date of the first sequence in either VP4/VP2 or VP1 (Table 2).

(e) The remaining genetic variants of HRV-C for which only VP4/VP2 region sequences are available and which show >10 % divergence from other species C sequences in this region should be assigned as provisionally assigned types (designated 'pat'), e.g. HRV-C_pat1, HRV-C_pat2 etc. (Table 3). If and when VP1 sequence data are determined for at least one member of this provisionally assigned type, it can be added to the list of confirmed types and removed from the provisional list.

(f) A designated Expert Group takes responsibility for the future coordinated assignment of HRV-C types, including a reappraisal of the type assignment as more sequence data accumulate. This might perhaps be nominated by the ICTV Picornavirus Study Group and include Study

Table 3. Provisional type assignment of HRV-C variants represented by VP4/VP2 sequences only

Provisional type	GenBank accession no.	Submission date	Variants	Reference*
HRV-C_pat1	EF077256	20 Oct 2006	21	Kistler *et al.* (2007)
HRV-C_pat2	EF077260	20 Oct 2006	8	Kistler *et al.* (2007)
HRV-C_pat3	EU081790	3 Aug 2007	15	Renwick *et al.* (2007)
HRV-C_pat4	EU081791	3 Aug 2007	13	Renwick *et al.* (2007)
HRV-C_pat5	EU081799	3 Aug 2007	20	Renwick *et al.* (2007)
HRV-C_pat6	EU081802	3 Aug 2007	29	Renwick *et al.* (2007)
HRV-C_pat7	EU081803	3 Aug 2007	7	Renwick *et al.* (2007)
HRV-C_pat8	EU081805	3 Aug 2007	11	Renwick *et al.* (2007)
HRV-C_pat9	EU081807	3 Aug 2007	13	Renwick *et al.* (2007)
HRV-C_pat10	EU590054	25 Mar 2008	14	Savolainen-Kopra *et al.* (2009a)
HRV-C_pat11	EU590061	25 Mar 2008	5	Savolainen-Kopra *et al.* (2009a)
HRV-C_pat12	EU590064	25 Mar 2008	5	Savolainen-Kopra *et al.* (2009a)
HRV-C_pat13	EU697839	5 May 2008	6	Briese *et al.* (2008)
HRV-C_pat14	EU697852	5 May 2008	1	Briese *et al.* (2008)
HRV-C_pat15	EU743925	22 May 2008	10	Dominguez *et al.* (2008)
HRV-C_pat16	EU752358	26 May 2008	2	Miller *et al.* (2009a)
HRV-C_pat17	EU752398	26 May 2008	6	Miller *et al.* (2009a)
HRV-C_pat18	EU752412	26 May 2008	21	Miller *et al.* (2009a)
HRV-C_pat19	FJ598096	29 Dec 2008	5	Currently unpublished
HRV-C_pat20	FJ615722	9 Jan 2009	1	Miller *et al.* (2009b)
HRV-C_pat21	FJ615737	9 Jan 2009	4	Miller *et al.* (2009b)
HRV-C_pat22	FJ615745	9 Jan 2009	1	Miller *et al.* (2009b)
HRV-C_pat23	FJ841957	17 Mar 2009	2	Calvo *et al.* (2009)
HRV-C_pat24	FJ869923	27 Mar 2009	1	Calvo *et al.* (2009)
HRV-C_pat25	FJ869950	27 Mar 2009	4	Calvo *et al.* (2009)
HRV-C_pat26	GQ466482	7 Aug 2009	1	Savolainen-Kopra *et al.* (2009c)
HRV-C_pat27	GU214340	18 Nov 2009	5	Piralla *et al.* (2009)
HRV-C_pat28	HM347248	24 May 2010	1	Currently unpublished

*Reference for first submitted sequence for each type.

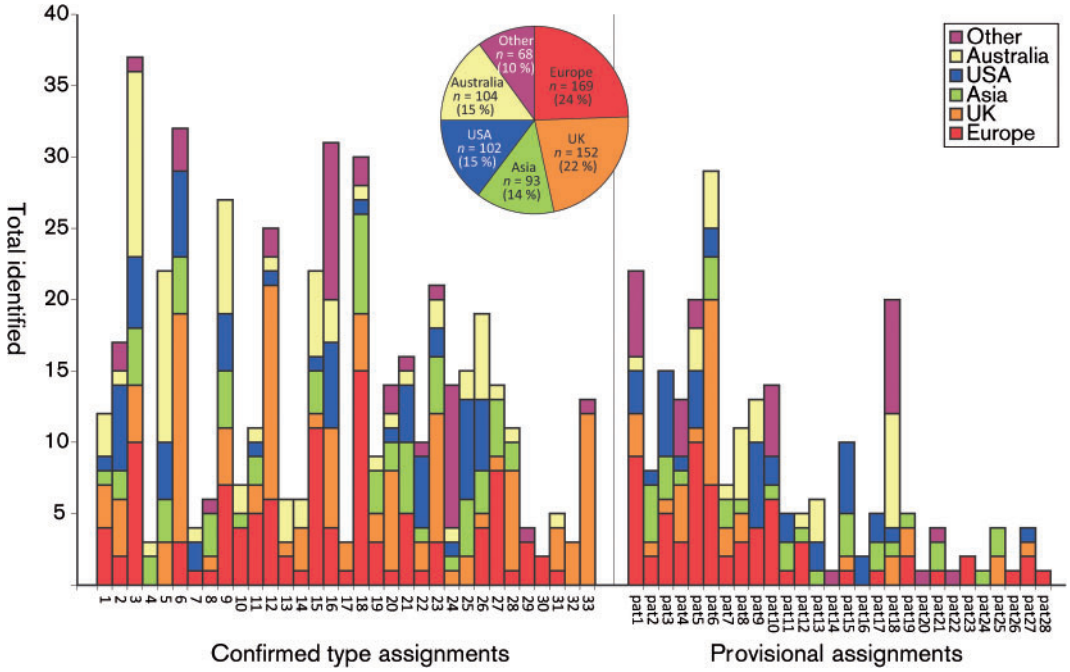


Fig. 5. Total numbers of each assigned HRV-C type identified by sequence comparisons in the VP4/VP2 region, divided by geographical region. The total representation of sequences from each geographical range is indicated by the inset pie chart.

Group members with expertise and experience in new enterovirus type assignments, as well as 'outside' scientists active in HRV-C or more general HRV research.

(g) Alignments of the VP1 and VP4/VP2 regions, along with information on the regions used for sequence comparisons, will be made available on a publicly available database accessible through the Picornavirus Study Group. These alignments will be regularly maintained and updated with new sequence data and type assignments as these become available.

(h) This Group should cooperate closely with those developing future type-assignment criteria for species A and B rhinoviruses to help ensure consistency in approach.

Applying these criteria to the currently available dataset of HRV-C sequences creates a total of 33 confirmed types (Tables 1 and 2) and a further 28 provisionally assigned types based on VP4/VP2 sequences (Table 3).

Type identification

In drawing up specific classification proposals, we should emphasize that the process of type assignment is an activity distinct from type identification or detection. From the data obtained from genetic characterization of HRV-C in different genome regions and the lack of recombination observed (McIntyre *et al.*, 2010), we consider that identification of an HRV-C genetic variant as belonging to an already classified type can be achieved by sequence comparisons in VP4/VP2. Currently, most sequence data obtained for genetic characterization of HRV-C have been obtained in this region, including all of the confirmed types. These sequence data are derived from a wide geographical base, combining sequence data from Europe, USA, Australia, Japan and South-East Asia.

By phylogenetic analysis and examination of pairwise distances within the now-extensive dataset of VP4/VP2 sequences, the aforementioned threshold permits all HRV-C variants characterized to date to be categorized into a total of 61 confirmed or provisionally assigned types, the majority of which now contain multiple examples from geographically separate locations (Fig. 5; a full list of individual assignments is available as Supplementary Table S1 in JGV Online). The decreasing pace of identification of variants worldwide that can be assigned (even provisionally) as new types suggests a finite limit to the number that will eventually be classified. The actual total will, however, only become clear with more temporally and geographically widespread sampling.

In summary, this proposal draws together existing knowledge of the genetic diversity of HRV-C and applies principles established for type assignment of other enteroviruses to create a genotypically based classification scheme for HRV-C types. We hope that these proposals will be of value in future rhinovirus research, and provide the impetus to develop related type-assignment criteria for novel HRV-A and -B genetic variants that have been described.

Acknowledgements

The authors would like to thank Nick Knowles of the Picornavirus Study Group for useful comments and advice received during the preparation of this proposal.

References

Arden, K. E. & Mackay, I. M. (2010). Newly identified human rhinoviruses: molecular methods heat up the cold viruses. *Rev Med Virol* **20**, 156–176.

Arden, K. E., McErlean, P., Nissen, M. D., Sloots, T. P. & Mackay, I. M. (2006). Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections. *J Med Virol* **78**, 1232–1240.

Arden, K. E., Chang, A. B., Lambert, S. B., Nissen, M. D., Sloots, T. P. & Mackay, I. M. (2010a). Newly identified respiratory viruses in children with non-hospitalised asthma exacerbation. *J Med Virol* **82**, 1458–1461.

Arden, K. E., Faux, C. E., O'Neill, N. T., McErlean, P., Nitsche, A., Lambert, S. B., Nissen, M. D., Sloots, T. P. & Mackay, I. M. (2010b). Molecular characterization and distinguishing features of a novel human rhinovirus (HRV) C, HRVC-QCE, detected in children with fever, cough and wheeze during 2003. *J Clin Virol* **47**, 219–223.

Blomqvist, S., Savolainen, C., Raman, L., Roivainen, M. & Hovi, T. (2002). Human rhinovirus 87 and enterovirus 68 represent a unique serotype with rhinovirus and enterovirus features. *J Clin Microbiol* **40**, 4218–4223.

Briese, T., Renwick, N., Venter, M., Jarman, R. G., Ghosh, D., Kondgen, S., Shrestha, S. K., Hoegh, A. M., Casas, I. & other authors (2008). Global distribution of novel rhinovirus genotype. *Emerg Infect Dis* **14**, 944–947.

Calvo, C., Luz Garcia, M., Pozo, F., Reyes, N., Perez-Brena, P. & Casas, I. (2009). Role of rhinovirus C in apparently life-threatening events in infants, Spain. *Emerg Infect Dis* **15**, 1506–1508.

Carstens, E. B. (2010). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2009). *Arch Virol* **155**, 133–146.

Cooney, M. K., Fox, J. P. & Kenny, G. E. (1982). Antigenic groupings of 90 rhinovirus serotypes. *Infect Immun* **37**, 642–647.

Cordey, S., Gerlach, D., Junier, T., Zdobnov, E. M., Kaiser, L. & Tapparel, C. (2008). The *cis*-acting replication elements define human enterovirus and rhinovirus species. *RNA* **14**, 1568–1578.

Dominguez, S. R., Briese, T., Palacios, G., Hui, J., Villari, J., Kapoor, V., Tokarz, R., Glode, M. P., Anderson, M. S. & other authors (2008). Multiplex MassTag-PCR for respiratory pathogens in pediatric nasopharyngeal washes negative by conventional diagnostic testing shows a high prevalence of viruses belonging to a newly recognized rhinovirus clade. *J Clin Virol* **43**, 219–222.

Gerber, K., Wimmer, E. & Paul, A. V. (2001). Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a *cis*-replicating element in the coding sequence of 2A(pro). *J Virol* **75**, 10979–10990.

Hamparian, V. V., Colonno, R. J., Cooney, M. K., Dick, E. C., Gwaltney, J. M., Jr, Hughes, J. H., Jordan, W. S., Jr, Kapikian, A. Z., Mogabgab, W. J. & Monto, A. (1987). A collaborative report: rhinoviruses – extension of the numbering system from 89 to 100. *Virology* **159**, 191–192.

Han, T. H., Chung, J. Y., Hwang, E. S. & Koo, J. W. (2009). Detection of human rhinovirus C in children with acute lower respiratory tract infections in South Korea. *Arch Virol* **154**, 987–991.

Huang, T., Wang, W., Bessaud, M., Ren, P., Sheng, J., Yan, H., Zhang, J., Lin, X., Wang, Y. & other authors (2009). Evidence of recombination and genetic diversity in human rhinoviruses in children with acute respiratory infection. *PLoS ONE* **4**, e6355.

Kaiser, L., Aubert, J. D., Pache, J. C., Deffernez, C., Rochat, T., Garbino, J., Wunderli, W., Meylan, P., Yerly, S. & other authors (2006). Chronic rhinoviral infection in lung transplant recipients. *Am J Respir Crit Care Med* **174**, 1392–1399.

Kapikian, A. Z., Conant, R. M., Hamparian, V. V., Chanock, R. M., Chapple, P. J., Dick, E. C., Fenters, J. D., Gwaltney, J. M., Jr, Hamre, D. & other authors (1967). Rhinoviruses: a numbering system. *Nature* **213**, 761–762.

Kapikian, A. Z., Conant, R. M., Hamparian, V. V., Chanock, R. M., Dick, E. C., Gwaltney, J. M., Jr, Hamre, D., Jordan, W. S., Kenny, G. E. & other authors (1971). A collaborative report: rhinoviruses – extension of the numbering system. *Virology* **43**, 524–526.

Khetsuriani, N., Lu, X., Teague, W. G., Kazerouni, N., Anderson, L. J. & Erdman, D. D. (2008). Novel human rhinoviruses and exacerbation of asthma in children. *Emerg Infect Dis* **14**, 1793–1796.

Kistler, A., Avila, P. C., Rouskin, S., Wang, D., Ward, T., Yagi, S., Schnurr, D., Ganem, D., DeRisi, J. L. & Boushey, H. A. (2007). Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* **196**, 817–825.

Knowles, N. J. (2010). Virus Taxonomy: 2009 release. http://www.ictvonline.org/virusTaxonomy.asp?version=2009

Laine, P., Savolainen, C., Blomqvist, S. & Hovi, T. (2005). Phylogenetic analysis of human rhinovirus capsid protein VP1 and 2A protease coding sequences confirms shared genus-like relationships with human enteroviruses. *J Gen Virol* **86**, 697–706.

Lamson, D., Renwick, N., Kapoor, V., Liu, Z., Palacios, G., Ju, J., Dean, A., St George, K., Briese, T. & Lipkin, W. I. (2006). MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004–2005. *J Infect Dis* **194**, 1398–1402.

Lau, S. K., Yip, C. C., Tsoi, H. W., Lee, R. A., So, L. Y., Lau, Y. L., Chan, K. H., Woo, P. C. & Yuen, K. Y. (2007). Clinical features and complete genome characterization of a distinct human rhinovirus (HRV) genetic cluster, probably representing a previously undetected HRV species, HRV-C, associated with acute respiratory illness in children. *J Clin Microbiol* **45**, 3655–3664.

Lau, S. K., Yip, C. C., Lin, A. W., Lee, R. A., So, L. Y., Lau, Y. L., Chan, K. H., Woo, P. C. & Yuen, K. Y. (2009). Clinical and molecular epidemiology of human rhinovirus C in children and adults in Hong Kong reveals a possible distinct human rhinovirus C subgroup. *J Infect Dis* **200**, 1096–1103.

Ledford, R. M., Patel, N. R., Demenczuk, T. M., Watanyar, A., Herbertz, T., Collett, M. S. & Pevear, D. C. (2004). VP1 sequencing of all human rhinovirus serotypes: insights into genus phylogeny and susceptibility to antiviral capsid-binding compounds. *J Virol* **78**, 3663–3674.

Lee, W. M., Kiesner, C., Pappas, T., Lee, I., Grindle, K., Jartti, T., Jakiela, B., Lemanske, R. F., Jr, Shult, P. A. & Gern, J. E. (2007). A diverse group of previously unrecognized human rhinoviruses are common causes of respiratory illnesses in infants. *PLoS ONE* **2**, e966.

McErlean, P., Shackelton, L. A., Lambert, S. B., Nissen, M. D., Sloots, T. P. & Mackay, I. M. (2007). Characterisation of a newly identified human rhinovirus, HRV-QPM, discovered in infants with bronchiolitis. *J Clin Virol* **39**, 67–75.

McIntyre, C., McWilliam-Leitch, E. C., Savolainen-Kopra, C., Hovi, T. & Simmonds, P. (2010). Analysis of genetic diversity and sites of

recombination in human rhinovirus species C (HRV-C). *J Virol* (in press). doi:10.1128/JVI.00962-10

Miller, E. K., Edwards, K. M., Weinberg, G. A., Iwane, M. K., Griffin, M. R., Hall, C. B., Zhu, Y., Szilagyi, P. G., Morin, L. L. & other authors (2009a). A novel group of rhinoviruses is associated with asthma hospitalizations. *J Allergy Clin Immunol* **123**, 98–104.

Miller, E. K., Khuri-Bulos, N., Williams, J. V., Shehabi, A. A., Faouri, S., Al Jundi, I., Chen, Q., Heil, L., Mohamed, Y. & other authors (2009b). Human rhinovirus C associated with wheezing in hospitalised children in the Middle East. *J Clin Virol* **46**, 85–89.

Mulders, M. N., Salminen, M., Kalkkinen, N. & Hovi, T. (2000). Molecular epidemiology of coxsackievirus B4 and disclosure of the correct VP1/2A(pro) cleavage site: evidence for high genomic diversity and long-term endemicity of distinct genotypes. *J Gen Virol* **81**, 803–812.

Oberste, M. S., Maher, K., Kilpatrick, D. R. & Pallansch, M. A. (1999). Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J Virol* **73**, 1941–1948.

Olenec, J. P., Kim, W. K., Lee, W. M., Vang, F., Pappas, T. E., Salazar, L. E., Evans, M. D., Bork, J., Roberg, K. & other authors (2010). Weekly monitoring of children with asthma for infections and illness during common cold seasons. *J Allergy Clin Immunol* **125**, 1001–1006.

Palmenberg, A. C., Spiro, D., Kuzmickas, R., Wang, S., Djikeng, A., Rathe, J. A., Fraser-Liggett, C. M. & Liggett, S. B. (2009). Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. *Science* **324**, 55–59.

Piotrowska, Z., Vazquez, M., Shapiro, E. D., Weibel, C., Ferguson, D., Landry, M. L. & Kahn, J. S. (2009). Rhinoviruses are a major cause of wheezing and hospitalization in children less than 2 years of age. *Pediatr Infect Dis J* **28**, 25–29.

Piralla, A., Rovida, F., Campanini, G., Rognoni, V., Marchi, A., Locatelli, F. & Gerna, G. (2009). Clinical severity and molecular typing of human rhinovirus C strains during a fall outbreak affecting hospitalized patients. *J Clin Virol* **45**, 311–317.

Renwick, N., Schweiger, B., Kapoor, V., Liu, Z., Villari, J., Bullmann, R., Miething, R., Briese, T. & Lipkin, W. I. (2007). A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. *J Infect Dis* **196**, 1754–1760.

Savolainen, C., Blomqvist, S., Mulders, M. N. & Hovi, T. (2002). Genetic clustering of all 102 human rhinovirus prototype strains: serotype 87 is close to human enterovirus 70. *J Gen Virol* **83**, 333–340.

Savolainen, C., Laine, P., Mulders, M. N. & Hovi, T. (2004). Sequence analysis of human rhinoviruses in the RNA-dependent RNA polymerase coding region reveals large within-species variation. *J Gen Virol* **85**, 2271–2277.

Savolainen-Kopra, C., Blomqvist, S., Kilpi, T., Roivainen, M. & Hovi, T. (2009a). Novel species of human rhinoviruses in acute otitis media. *Pediatr Infect Dis J* **28**, 59–61.

Savolainen-Kopra, C., Blomqvist, S., Smura, T., Roivainen, M., Hovi, T., Kiang, D., Kalra, I., Yagi, S., Louie, J. K. & other authors (2009b). 5′ noncoding region alone does not unequivocally determine genetic type of human rhinovirus strains. *J Clin Microbiol* **47**, 1278–1280.

Savolainen-Kopra, C., Blomqvist, S., Kaijalainen, S., Juonio, U., Juvonen, R., Peitso, A., Saukkoriipi, A., Vainio, O., Hovi, T. & Roivainen, M. (2009c). All known human rhinovirus species are present in sputum specimens of military recruits during respiratory infection. *Viruses* **1**, 1178–1189.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Tapparel, C., Junier, T., Gerlach, D., Van Belle, S., Turin, L., Cordey, S., Muhlemann, K., Regamey, N., Aubert, J. D. & other authors (2009a). New respiratory enterovirus and recombinant rhinoviruses among circulating picornaviruses. *Emerg Infect Dis* **15**, 719–726.

Tapparel, C., L'Huillier, A. G., Rougemont, A. L., Beghetti, M., Barazzone-Argiroffo, C. & Kaiser, L. (2009b). Pneumonia and pericarditis in a child with HRV-C infection: a case report. *J Clin Virol* **45**, 157–160.

Wisdom, A., Kutkowska, A., McWilliam-Leitch, E. C., Gaunt, E., Templeton, K., Harvala, H. & Simmonds, P. (2009a). Genetics, recombination and clinical features of human rhinovirus species C (HRV-C) infections; interactions of HRV-C with other respiratory viruses. *PLoS ONE* **4**, e8518.

Wisdom, A., McWilliam Leitch, C., Gaunt, E., Harvala, H. & Simmonds, P. (2009b). Screening respiratory samples for human rhinoviruses (HRV) and enteroviruses: comprehensive VP4/2-typing reveals high incidence and genetic diversity of HRV species C. *J Clin Microbiol* **47**, 3958–3967.

Journal of Medical Virology 84:536–542 (2012)

High Detection Frequency and Viral Loads of Human Rhinovirus Species A to C in Fecal Samples; Diagnostic and Clinical Implications

H. Harvala,[1]* C.L. McIntyre,[2] N.J. McLeish,[2] J. Kondracka,[2] J. Palmer,[2] P. Molyneaux,[3] R. Gunson,[4] S. Bennett,[4] K. Templeton,[1] and P. Simmonds[2]

[1]*Specialist Virology Centre, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom*
[2]*Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh*
[3]*Department of Medical Microbiology, Aberdeen Royal Infirmary, Aberdeen, United Kingdom*
[4]*West of Scotland Specialist Virology Centre, Gartnavel General Hospital, Glasgow, United Kingdom*

Human rhinoviruses (HRVs) can be divided into three species; HRV-A to HRV-C. Up to 148 different HRV (sero)types have been identified to date. Because of sequence similarity between 5′-NCR of HRVs and enteroviruses (EVs), it is problematic to design EV-specific RT-PCR assays. The aims of this study were to assess the rate of false-detection of different rhinoviruses by EV RT-PCR, and to evaluate the diagnostic and clinical significance of such cross-reactivity. In vitro RNA transcripts of HRV A-C created from cDNA templates were quantified spectrophotometrically. Six hundred twenty-one stool samples screened as part of routine diagnostic for EV, 17 EV-positive stool samples referred for typing, 288 stool samples submitted for gastroenteritis investigations, and 1,500 CSF samples were included in the study. EV-specific RT-PCR detected RNA transcripts of HRV-A1b, HRV-B14, and HRV-Crpat18 but with 10–1,000 reduced sensitivity compared to EV transcripts. Screening fecal samples by EV RT-PCR identified 13 positive samples identified subsequently as rhinoviruses; a further 26 HRV-positive samples were identified by nested HRV RT-PCR. All individuals were hospitalized and presented mostly with diarrhea. A total of 26 HRV types were identified (HRV-A: 46%; HRV-B: 13%; HRV-C: 41%). Results confirm that EV-specific RT-PCR can detect HRVs, and at a practical level, identify potential problems of interpretation if fecal samples are used for surrogate screening in cases of suspected viral meningitis. High detection frequencies (10%) and viral loads in stool samples provide evidence for enteric replication of HRV, and its association with enteric disease requires further etiological studies. ***J. Med. Virol. 84:536–542, 2012.***

KEY WORDS: human rhinovirus; HRV-A to C; gastroenteritis; diagnostic; enterovirus

INTRODUCTION

Human rhinoviruses (HRVs) together with enteroviruses (EVs) belong to the genus *Enterovirus* within the family *Picornaviridae*. HRVs can be divided into three species; HRV species A (HRV-A) comprises of 74 different serotypes, HRV-B of 25 serotypes and more recently characterized HRV-C of at least 49 types [Stanway et al., 2005; McIntyre et al., 2010]. In most cases, HRV causes mild and self-limiting upper respiratory tract infections or common colds, but HRVs have also been associated with bronchiolitis, lower respiratory infection as well as with exacerbation of asthma and chronic obstructive airway diseases [reviewed by Mackay, 2008].

HRV infections are most often diagnosed from respiratory tract sample by reverse transcription polymerase chain reaction (RT-PCR) targeting the 5′-non-coding region (5′-NCR) of the viral genome [Scheltinga et al., 2005; Mackay, 2008]. Although relatively highly conserved compared to HRV coding regions, sequence variation in the 5′-NCR between different HRV (sero)types creates difficulties in designing primer pairs that can satisfactorily detect all HRV types and species [Faux et al., 2011]. Additionally, sequence similarity in this region

*Correspondence to: H. Harvala, PhD, MSc, MD, Specialist Virology Centre, Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh EH16 4SA, UK.
E-mail: heli.simmonds@hotmail.com

Accepted 15 November 2011
DOI 10.1002/jmv.23203
Published online in Wiley Online Library (wileyonlinelibrary.com).

between HRVs and enteroviruses (EVs) complicates the development of EV-specific RT-PCR screening methods that would not also detect some HRVs.

The current study has two aims. Firstly, the frequency of cross-reactivity and relative sensitivity of EV RT-PCR for HRV, both as pre-quantified RNA transcripts and HRV-positive clinical specimens were investigated. The second aim was to investigate the frequency of enteric excretion of rhinoviruses by HRV-specific PCR and virological and clinical characterization of positive subjects.

MATERIALS AND METHODS

In Vitro Transcribed RNA for Species A–C Rhinoviruses

Genomic cDNA clones of two different HRVs, HRV-A1b (species A) and HRV-B14 (species B) were kindly provided by Professor G. Stanway [HRV-A1b; Hughes et al., 1988 and HRV-B14; Alsaadi et al., 1989]. Full-length transcripts were generated from these using the Ambion T7 Transcription Kit (Promega (UK), Southampton, UK). In vitro transcribed RNA for species C rhinoviruses were made from cDNA clones of the combined 5′-NCR/VP4 region of two HRV-C variants directly amplified from clinical specimens (types HRV-CrPat19 and HRV-C40). Transcripts were purified with the RNeasy Kit (Qiagen (UK), Crawley, West Sussex, UK) and concentrations were determined by NanoDrop2000 (Thermo Scientific/Fisher, Horsham, West Sussex, UK). Integrity of RNA transcripts was demonstrated by denaturing agarose gel electrophoresis (data not shown). Ten-fold dilution serious of each RNA transcript were made and stabilized using citrate buffer (pH 6.0, Ambion, Life Technologies, Glasgow, UK) supplemented with 0.05 μg/ml carrier tRNA and 0.1 U/ml RNAsin inhibitor.

RT-PCR Assays and Sequencing

HRV and EV detection in a diagnostic laboratory used one-step RT-PCR assays as modified from published protocols [Scheltinga et al., 2005; Dierssen et al., 2008; Bennett et al., 2011], whereas the additional retrospective screening was done using the nested RT-PCR targeting the 5′-UTR as prescribed previously [Wisdom et al., 2009]. EV-positive samples were primarily typed by VP1 sequencing [Leitch et al., 2009]; those refractory to EV typing were amplified and sequenced in the VP4 region for type identification [Wisdom et al., 2009]. Similarly, HRV-positives sample were further identified by VP4 sequencing [Wisdom et al., 2009].

Clinical Samples

A total of 621 diagnostic stool samples predominantly referred from young children (age under 2 years; 63% of samples) with suspected viral meningitis or pyrexia of unknown origin during a 14-month period (March 2010 to April 2011) were investigated in this study. This study additionally included 17 EV-positive stool samples referred from Aberdeen for enterovirus typing and a subset of stool samples (n = 288) submitted mostly for gastroenteritis investigations (mostly suspected norovirus infection) in May (n = 90), September (n = 108), and December (n = 90). Finally, 1,500 CSF samples from the Specialist Virology Center were also included.

All samples tested retrospectively were anonymized before testing, but epidemiologic and demographic information was retained while patient confidentiality was protected (in accordance with a protocol approved by the Lothian Regional Ethics Committee [Protocol 2002/4/36]).

RESULTS

Detection of HRV Sequences by 5′-UTR PCR

The number of mismatches between EV RT-PCR primers and probes used for diagnostic screening with representative sequences of all assigned HRV and EV (sero)types was determined (Table I). This revealed a substantially greater number of mismatches between primers and probes with all three HRV species than with EV. Mismatches were similarly identified between the probe and primers sequences with the HRV RNA transcript sequences used in the current study as assay controls.

The actual degree of assay cross-reactivity was determined experimentally for HRV sequences by comparing C_t-values of HRV RNA transcripts derived from the cDNA clones of HRV-A1b, HRV-B14, HRV-CrPat19, and HRV-C40 amplified by both methods (Fig. 1). RNA transcripts were quantified biophysically before dilution for assay calibration into RNA copies. The four HRV transcripts showed different amplification kinetics in the HRV RT-PCR; the end-point titers for HRV-B14, HRV-C40, and HRV-CrPat19 transcripts were one, two, and three logs lower than for HRV-A1b. However, the EV-specific RT-PCR was also able to detect RNA transcripts of HRV-A1b, HRV-B14, and HRV-C19 but with substantially less sensitivity; 9,000 copies of HRV-A1b and HRV-B14, and 90,000 copies of HRV-C19 could be detected. Interestingly, the detection efficiency of HRV-A1b and HRV-B14 by EV-specific RT-PCR assay was comparable to that of HRV-C19 by HRV-specific RT-PCR.

EV RT-PCR Cross-Reactivity in Clinical Samples

A total of 67 EV RT-PCR reactive samples identified on diagnostic screening of 621 stool samples (frequency of 11%) were further characterized in this study. From those, seven samples untypeable by the EV-specific VP1 typing assays were subsequently

TABLE I. Number of Mismatches Between Primer and Probe Sequences of the HRV and EV RT-PCRs With Prototype HRV and EV 5′-UTR Sequences

	Mean number of mismatches (range)			
	Forward1	Forward2	Reverse	Probe
(A) Rhinovirus PCR				
Rhinovirus species				
HRV-A (75 types)	1.3 (0–2)	4.2 (3–5)	1.2 (1–3)	1 (1–2)
HRV-B (28 types)	1.8 (1–4)	2.2 (2–3)	1.6 (1–2)	1 (1–2)
HRV-C (10 types)	1.4 (1–3)	4 (4)	2 (1–3)	1 (1)

	Mean number of mismatches (range)		
	Forward	Reverse	Probe
(B) Enterovirus PCR			
Rhinovirus species			
HRV-A (75 types)	5.9 (3–7)	3.5 (1–4)	1.7 (0–3)
HRV-B (28 types)	5.5 (3–7)	2.4 (2–3)	1 (1–2)
HRV-C (10 types)	4.4 (3–7)	2.7 (1–4)	0.9 (0–2)
Enterovirus species			
EV-A (17 types)	1 (0–2)	0	0 (0–1)
EV-B (58 types)	1 (0–3)	0 (0–1)	0 (0–1)
EV-C (24 types)	2.3 (2–4)	0 (0–1)	0 (0–2)
EV-D (3 types)	2 (1–3)	0	0 (0–1)

identified as HRV-positives by VP4 amplification and sequencing (10.4%; Table II). VP4 typing similarly identified 6 HRV-positive samples in a second group of 17 EV-positive stool samples referred from Aberdeen (35%; Table II).

To investigate the frequency of HRV excretion in feces, 288 stool samples were retrospectively screened for HRV by nested RT-PCR. A total of 26 further HRV-positive samples were further identified and typed by VP4 amplification and sequencing (frequency 10%, including three identified by EV RT-PCR, Table III). Detection frequencies varied from 7.4% in May and 7.7% in December to 12.2% in September. One-step real-time RT-PCR would have detected only 11 of these 26 samples (42%; data not shown).

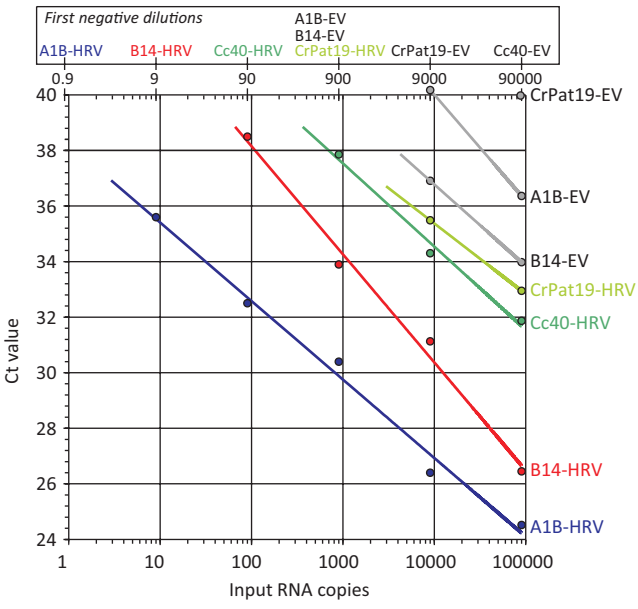


Fig. 1. Amplification of RNA transcripts of defined concentrations by EV RT-PCR and HRV RT-PCR. C_t-values (y-axis) were plotted for serial 10-fold dilutions of each transcript (x-axis). The PCR assay used for amplification (EV, HRV) is indicated as a suffix to each label (e.g., HRV-1B–EV: amplification of the HRV-1B transcript in the EV RT PCR).

Viral loads of HRV and EV in fecal samples were comparable (Fig. 2). Mean C_t-values of the HRV-positive samples in the HRV RT-PCR and of EV-positive samples in the EV RT-PCR were 28.5 (standard deviation ±3.5) and 28.9 (±1.9), respectively ($P = 0.77$).

No specific type of HRV was associated with detection in fecal samples. A total of 26 different HRV types accounted for the 39 identified HRV-positive samples; 16 samples (41%) were identified as species A (HRV-A1, A7, A19, A34, A40, A46, A71, A78, A89, and A103), 5 (13%) as species B (HRV-B6, B52, and B97), and 18 (46%) as species C (HRV-C5, C6, C12, C14, C24, C25, C36, C37, C39, C42, C43, and C45). Positive samples were mostly obtained from young children under the age of 2 years (26/39) or from the patients over the age of 65 years (7/39). All of these individuals were hospitalized, and seven of them required intensive therapy unit (ITU) admission. Patients presented with variety of clinical symptoms, most often with fever, diarrhea, and/or vomiting; 72% of patients did not have any respiratory symptoms at admission (28/39). Nine patients were investigated as part of unconfirmed norovirus outbreak in hospital. A total of 11 patients (28%) presented with acute respiratory symptoms, accompanied with loose stool or diarrhea. Enteric co-pathogens were identified in

TABLE II. HRV-Positive Patients Identified by Real-Time EV RT-PCR

Date	Age	Clinical details	EV PCR	HRV PCR	Type
Nov_10	7–12 months	Bronchiolitis, apnoeic episodes	38.09	32.89	HRV-B52
Nov_10	7–12 months	Fever, diarrhea and vomiting	33.89	30.78	HRV-C25
Feb_11	3–5 years	LRTI and fever	41.1	ND	HRV-C6
May_10	1–2 years	URTI, fever and diarrhea	38.48	ND	HRV-A40[a]
July_10	3–6 months	Seizure and loose stool	36.31	ND	HRV-A46
Sep_10	7–12 months	Bronchiolitis and diarrhea	36.46	ND	HRV-C12
Sep_10	<3 months	Loose stool and poor weight gain	42.55	29.01	HRV-C42
Aug_10	3–6 months	Loose stool and fever	NA	28.55	HRV-B52
Aug_10	6–12 months	Diarrhea	NA	26.39	HRV-A78
Aug_10	<3 months	Fever and poor feeding	NA	29.62	HRV-B6[b]
Aug_10	>65 years	Seizure and loose stool	NA	30.39	HRV-A7
Sep_10	10–15 years	Fever, diarrhea and vomiting	28.02	23.43	HRV-A71
Nov_10	1–2 years	LRTI and fever	33.01	28.75	HRV-C5

NA, data not available; ND, not done, sample finished.
[a]One co-infection with adenovirus identified.
[b]One co-infection with rotavirus identified.

five patients (adenovirus n = 3, rotavirus n = 1, and *Salmonella gastroenteritidis* n = 1).

To investigate whether non-specificity of EV RT-PCR extended to diagnostic testing of CSF samples, and indeed whether HRV can cause CNS-associated infection, 1,500 CSF samples were screened by nested HRV RT-PCR. From these, HRV RNA was detected in a total of one individual sample; this sample was negative by routine EV RT-PCR assay most likely due to a low viral load. A finding was confirmed after re-extraction of RNA from original CSF sample. The sample was obtained from a young girl under the age of 3 months who presented with suspected sepsis-like illness, and has also HRV detected in her throat swab. The virus was identified as HRV-B6 in both CSF sample and throat swab (stool sample not obtained).

DISCUSSION

This study documents substantial cross-reactivity between EV and HRV real-time PCR methods for HRV RNA sequences and identifies a relatively high

TABLE III. Characteristics of Patients Identified as HRV-Positive by Screening Stool Samples by HRV-Specific RT-PCR Assay

Sample	Date	Age	Location	Clinical details	Type
10_91	Sep_10	>65 years	RIE	Diarrhea[a]	HRV-A1
10_92	Sep_10	37–65 years	ITU	Diarrhea and fever[a]	HRV-A1
10_93	Sep_10	>65 years	ITU	Diarrhea and fever[a]	HRV-A1
10_94	Sep_10	1–2 years	RHSC	Diarrhea and fever	HRV-A1
10_95	Sep_10	<3 months	RHSC	Diarrhea and vomiting	HRV-A1
10_100	Sep_10	7–12 months	RHSC	Loose stool and irritable	HRV-C24
10_128	Sep_10	<3 months	RHSC	Bronchiolitis, loose stool	HRV-C43
10_129	Sep_10	1–2 years	RHSC	Fever (post-transplant)	HRV-C43[b]
10_141	Sep_10	3–6 months	RHSC	Diarrhea	HRV-C39[c]
10_149	Sep_10	2 years	RHSC	URTI and loose stool	HRV-A19[b]
10_153	Sep_10	>65 years	ITU	Diarrhea and fever[a]	HRV-C39
10_511	May_10	6–10 years	RHSC	Diarrhea	HRV-C45
10_512	May_10	7–12 months	RHSC	URTI, diarrhea and vomiting	HRV-C12
10_518	May_10	3–5 years	RHSC	Diarrhea and fever	HRV-A34
10_520	May_10	7–12 months	ESCITU	Diarrhea and seizure	HRV-A89
10_521	May_10	<3 months	ESCITU	Diarrhea, vomiting and fever	HRV-C37
10_522	May_10	<3 months	RHSC	Bronchiolitis and vomiting	HRV-C36
10_547	May_10	7–12 months	RHSC	Diarrhea and vomiting[a]	HRV-C36
10_568	May_10	3–5 years	RHSC	Diarrhea and vomiting[a]	HRV-A103
10_121	Dec_10	3–6 months	ESCITU	Bronchiolitis and vomiting	HRV-A34
10_123	Dec_10	7–12 months	ESCITU	Bronchiolitis and vomiting	HRV-A89
10_124	Dec_10	7–12 months	RHSC	Gastroenteritis	HRV-C37
10_125	Dec_10	7–12 months	RHSC	Diarrhea and seizure	HRV-C36
10_128	Dec_10	>65 years	RIE	Diarrhea and vomiting[a]	HRV-B97
10_1213	Dec_10	>65 years	RIE	Diarrhea and vomiting[a]	HRV-C14
10_1218	Dec_10	>65 years	RIE	Diarrhea and vomiting[a]	HRV-B52

RIE, Royal Infirmary Edinburgh; RHSC, Royal Hospital for Sick Children; ITU, Intensive Therapy Unit; ESCITU, Edinburgh Sick Children ITU.
[a]Investigated as part of norovirus outbreak.
[b]Two co-infections with adenovirus identified.
[c]One co-infection with *Salmonella enteritidis* identified.

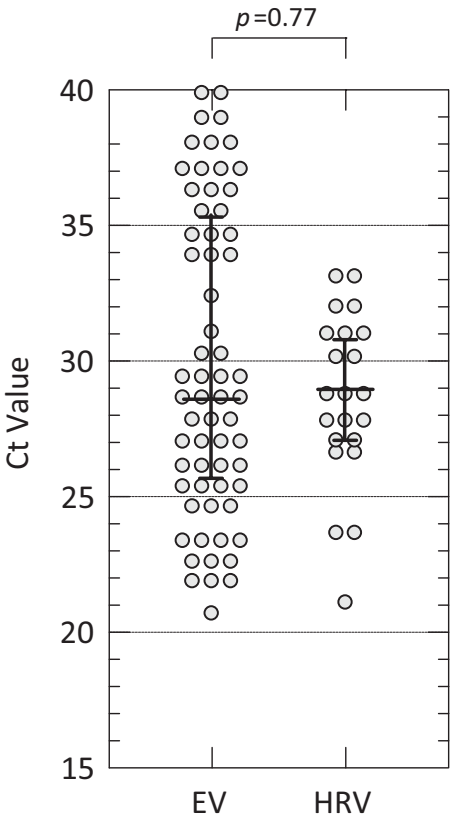


Fig. 2. C_t-values of EV-positive samples in the EV RT-PCR (left graph) and HRV-positive samples by the HRV RT-PCR (right graph). Mean values and 67 percentile distributions shown as bars. C_t-values were compared by the non-parametric Kruskall–Wallace test, *P*-value shown above graph.

frequency of HRV excretion in feces (comparable to EV detection rates). The similarity in viral loads between HRV and EV in feces provides evidence for a much greater propensity of HRV to establish enteric infections than previously described.

Detection of HRV by EV RT-PCR

Although infrequently described and not formally quantified [Faux et al., 2011], cross-reactivity of HRV in EV detection assays may be reasonably anticipated given the sequence similarity of the region of the 5′-UTR typically targeted by EV screening assays. Both HRVs and EVs have a type I internal ribosomal entry site and there are substantial similarities in key functional areas including RNA secondary structures between all members of the *Enterovirus* genus that extend to the targeted region (nucleotides from 412 to 553). Indeed, the greater number of mismatches between the EV RT-PCR primers and probes with the HRV-A1b and HRV-B14 than with EV sequences (max 6, 3, and 1 for sense primer, antisense primer, and probe) did not preclude their amplification by HRV RT-PCR, although with a sensitivity approximately 100-fold lower that for EV sequences.

The detection frequency and potential mis-identification of HRV-positive samples as EV positive using standard EV RT-PCR creates a number of interpretation difficulties for diagnostic screening of fecal samples and to a much more limited extent in the current study of CSF samples. The relatively high viral loads of HRV in fecal samples potentially creates problems for laboratories that use this sample type as an additional or alternative surrogate sample to CSF samples to guide diagnosis of EV meningitis and encephalitis as frequently directed by local or national guidelines. As shown here, approximately 10% of stool samples referred for virology diagnostic testing are genuine HRV-positive generally with high viral loads. If even a small proportion of these cross-reacted in the EV RT-PCR, as can be predicted from the study findings where approximately 15% (13/84) of EVs identified by EV RT-PCR in stool samples were indeed HRVs, then this could create a number of falsely reported EV detections in meningitis cases caused by other organisms. In this situation, virus typing as an adjunct to screening that included methods to detect HRV variants would identify instances of false-detection of EVs; our analysis of CSF samples does not suggest HRV being a frequent cause of central nervous system associated infections.

Tissue Tropism of HRV

An unanticipated finding from the current study was the high rate of HRV detection in fecal samples and the comparability of their viral loads to enteroviruses that are long known to replicate primarily in the lower gastrointestinal tract [Sabin and Ward, 1941; Iwasaki et al., 2002]. HRV are labile to the acidic environment of the stomach that leads to irreversible conformational changes in capsid. Experimentally, HRVs differ from enteroviruses in being inactivated at pH of <5 as well as pH values above 9–10. For this reason, HRV detection in fecal samples or sewage has generally been attributed in the past to swallowed respiratory secretions [Blomqvist et al., 2009]. In our opinion, this is unlikely for two reasons; firstly, the acid lability of HRVs that destroys their infectivity when passing through the stomach would lead to HRV RNA degradation if the virion capsid was breached. Secondly, the dilution of HRV secretions with the much larger volumes of gut contents could not conceivably create the high viral loads detected in the current study by HRV PCR. Indeed, the simplest explanation for the near identity of viral loads of HRV and EV in fecal samples is that viruses replicate comparably in the gastrointestinal tract. However, studies are required to demonstrate this further. The restriction of HRV-positive samples to children of <2 years of age and to the elderly may have arisen from survival of infectious HRV through the less acidic stomach environment in these groups [Lopez-Alonso and Ribas, 1991]. Survival of HRV may additionally be enhanced by the presence of cationic substances

such as $MgCl_2$ that are favorable for the replication of HRV by stabilizing virions [Blough et al., 1969]. As an additional possibility, in vitro culture of HRV-B14 in acidic medium has been shown to select rapidly for a point mutation in the capsid (VP1) gene that confers resistance to low pH [Skern et al., 1991]. This may represent a more general and necessary adaptive change for successful replication in the GI tract. Whether each of the wide range of HRV (sero)types and species detected in fecal samples in the current study are capable of an equivalent phenotypic change and identification of the associated genetic changes are intriguing areas of further research. The other potential requirement for successful gut replication by rhinoviruses is adaptation to replication at 37°C instead of the 33–34°C in the respiratory tract, traditionally assumed to be the optimum temperature of HRVs. However, when actually experimentally investigated [Papadopoulos et al., 1999], a wide range of reference species A and B serotypes and primary clinical isolates showed minimal differences in replication ability at 33 and 37°C, and indeed some isolates actually grew better at the higher incubation temperature.

The frequent detection and high viral loads of HRV in fecal samples in the current study is consistent with the frequent detection of HRV by virus isolation during environmental screening of human sewage [Blomqvist et al., 2009]. It was speculated that the HRV variants derived from respiratory secretions carried into sewage by discarded used tissues or in wastewater. Our findings of potential gut replication of HRV implied by the high viral loads in fecal samples (Fig. 2) provide a more direct explanation for the presence of infectious HRV in sewage. The plethora of different HRV types (n = 26) detected in fecal samples argues against the existence of specifically enteropathic strains of HRV. Furthermore, using PCR-based screening methods equally sensitive for all three species, the relative proportions of HRV-A (46%), HRV-B (13%), and HRV-C (41%) were broadly comparable to previously reported detection frequencies in respiratory samples [63%, 7%, and 30%, respectively; Wisdom et al., 2009].

HRV have traditionally been associated with respiratory tract infections. Although HRV are mostly regarded as harmless cause of common colds, they have been linked more recently with severe neonatal infections [van Piggelen et al., 2010; Broberg et al., 2011] and pericarditis [Tapparel et al., 2009]. In these case reports HRV was also detected in stool samples, consistent with the more severe systemic nature of HRV infections in young children. Although not formally tested by prospective study and without control groups, our data support the idea that HRV may on some occasions be responsible cause of gastroenteritis, given the high frequency of diarrhea, absence of any respiratory symptoms in 70% of the positive subjects as well as absence of detected co-pathogens. Indeed, eight of our patients were investigated as a part of suspected gastroenteritis outbreak in hospital, and HRV (in fecal samples) has been the only etiological agent identified so far. More focused screening of fecal samples for HRV in patients presenting with gastroenteritis, sepsis, and other systemic diseases, along with parallel estimation of HRV detection frequency in control groups will be required to demonstrate these additional potential disease association of HRV infections. Furthermore, the possible presence of infectious HRV in fecal samples potentially creates infection control problems. Whereas HRV are usually regarded to spread via direct contact with respiratory droplets, findings in the current study point towards the very real possibility of fecal–oral transmission of HRV. This would have a number of implications for the management of HRV-infected patients in a hospital or institutional setting.

REFERENCES

Alsaadi S, Hassard S, Stanway G. 1989. Sequences in the 5′ noncoding region of human rhinovirus 14 RNA that affect in vitro translation. J Gen Virol 70:2799–2804.

Bennett S, Harvala H, Witteveldt J, McWilliam-Leitch EC, McLeish N, Templeton K, Gunson R, Carman WF, Simmonds P. 2011. Rapid simultaneous detection of enterovirus and parechovirus RNA in clinical samples by one-step real-time RT-PCR assay. J Clin Microbiol 49:2620–2624.

Blomqvist S, Savolainen-Kopra C, Paananen A, Hovi T, Roivainen M. 2009. Molecular characterization of human rhinovirus field strains isolated during surveillance of enteroviruses. J Gen Virol 30:1371–1381.

Blough HA, Tiffany JM, Gordon G, Fiala M. 1969. The effect of magnesium on the intracellular crystallization of rhinovirus. Virology 38:694–698.

Broberg E, Niemela J, Lahti E, Hyypia T, Ruuskanen O, Waris M. 2011. Human rhinovirus C–associated severe pneumonia in a neonate. J Clin Virol 51:79–82.

Dierssen U, Rehren F, Henke-Gendo C, Harste G, Heim A. 2008. Rapid routine detection of enterovirus RNA in cerebrospinal fluid by a one-step real-time RT-PCR assay. J Clin Virol 42:58–64.

Faux CE, Arden KE, Lambert SB, Nissen MD, Nolan TM, Chang AB, Sloots TP, Mackay IM. 2011. Usefulness of published PCR primers in detecting human rhinovirus infection. Emerg Infect Dis 17:296–298.

Hughes PJ, North C, Jellis CH, Minor PD, Stanway G. 1988. The nucleotide sequence of human rhinovirus 1B: Molecular relationships within the rhinovirus genus. J Gen Virol 69:49–58.

Iwasaki A, Welker R, Mueller S, Linehan M, Nomoto A, Wimmer E. 2002. Immunofluorescence analysis of poliovirus receptor expression in Peyer's patches of humans, primates, and CD155 transgenic mice: Implications for poliovirus infection. J Infect Dis 186:585–592.

Leitch EC, Harvala H, Robertson I, Ubillos I, Templeton K, Simmonds P. 2009. Direct identification of human enterovirus serotypes in cerebrospinal fluid by amplification and sequencing of the VP1 region. J Clin Virol 44:119–124.

Lopez-Alonso M, Ribas J. 1991. Technical improvement for anorectal manometry in newborns. J Pediatr Surg 26:1215–1218.

Mackay IM. 2008. Human rhinoviruses: The cold wars resume. J Clin Virol 42:297–320.

McIntyre CJ, McWilliam-Leitch EC, Savolainen-Kopra C, Hovi T, Simmonds P. 2010. Analysis of genetic diversity and sites of recombination in human rhinovirus species C (HRV-C). J Virol 83:10297–10310.

Papadopoulos NG, Sanderson G, Hunter J, Johnston SL. 1999. Rhinoviruses replicate effectively at lower airway temperatures. J Med Virol 58:100–104.

Sabin AB, Ward RR. 1941. The natural history of human poliomyelitis. J Exp Med 73:771–793.

Scheltinga SA, Templeton KE, Beersma MF, Claas EC. 2005. Diagnosis of human metapneumovirus and rhinovirus in patients with respiratory tract infections by an internally controlled multiplex real-time RNA PCR. J Clin Virol 33:306–311.

Skern T, Torgersen H, Auer H, Kuechler E, Blaas D. 1991. Human rhinovirus mutants resistant to low pH. Virology 183:757–763.

Stanway G, Brown F, Christian P, Hovi T, Hyypia T, King AMQ, Knowles NJ, Lemon SM, Minor PD, Pallansch MA, Palmenberg AC, Skern T. 2005. Family picornaviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger V, Ball La, editors. Virus taxonomy. Eight report on the International Committee on Taxonomy of Viruses. London: Elsevier/Academic Press. pp 757–778.

Tapparel C, L'Huillier AG, Rougemont AL, Beghetti M, Barazzone-Argiroffo C, Kaiser L. 2009. Pneumonia and pericarditis in a child with HRV-C infection: A case report. J Clin Virol 45:157–160.

van Piggelen RO, Van Loon AM, Krediet TG, Verboon-Maciolek MA. 2010. Human rhinovirus causes severe infection in preterm infants. Pediatr Infect Dis J 29:364–365.

Wisdom A, McWilliam Leitch C, Gaunt E, Harvala H, Simmonds P. 2009. Screening respiratory samples for human rhinoviruses (HRV) and enteroviruses: Comprehensive VP4/2-typing reveals high incidence and genetic diversity of HRV species C. J Clin Microbiol 47:3958–3967.